

SCHEME : K

Name : _____
Roll No. : _____ Year : 20__ 20__
Exam Seat No. : _____

LABORATORY MANUAL FOR DATA STORY TELLING AND VISUALIZATION (313004)



COMPUTER ENGINEERING GROUP



**MAHARASHTRA STATE BOARD OF
TECHNICAL EDUCATION, MUMBAI
(Autonomous) (ISO 9001: 2015) (ISO/IEC 27001:2013)**

Vision

To ensure that the Diploma Level Technical Education constantly matches the latest requirements of technology industry and includes the all-round personal development of students including social concerns and to become globally competitive, technology led organization.

Mission

To provide high quality technical and managerial manpower, information and consultancy services to the industry and community to enable the industry and community to face the changing technological & environmental challenges.

Quality Policy

We, at MSBTE are committed to offer the best in class academic services to the students and institutes to enhance the delight of industry and society. This will be achieved through continual improvement in management practices adopted in the process of curriculum design, development, implementation, evaluation and monitoring system along with adequate faculty development programs.

Core Values

MSBTE believes in the following:

- Education industry produces live products,
- Market requirements do not wait for curriculum changes.
- Question paper is the reflector of academic standards of educational organization.
- Well-designed curriculum needs effective implementation too.
- Competency based curriculum is the backbone of need based program.
- Technical skills do need support of life skills,
- Best teachers are the national assets.
- Effective teaching learning process is impossible without learning resources.

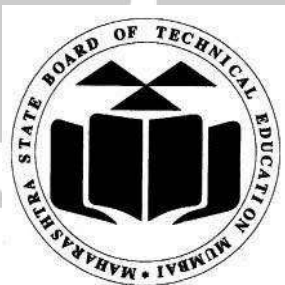
A Laboratory Manual for

**DATA STORY TELLING AND
VISUALIZATION
(313004)**

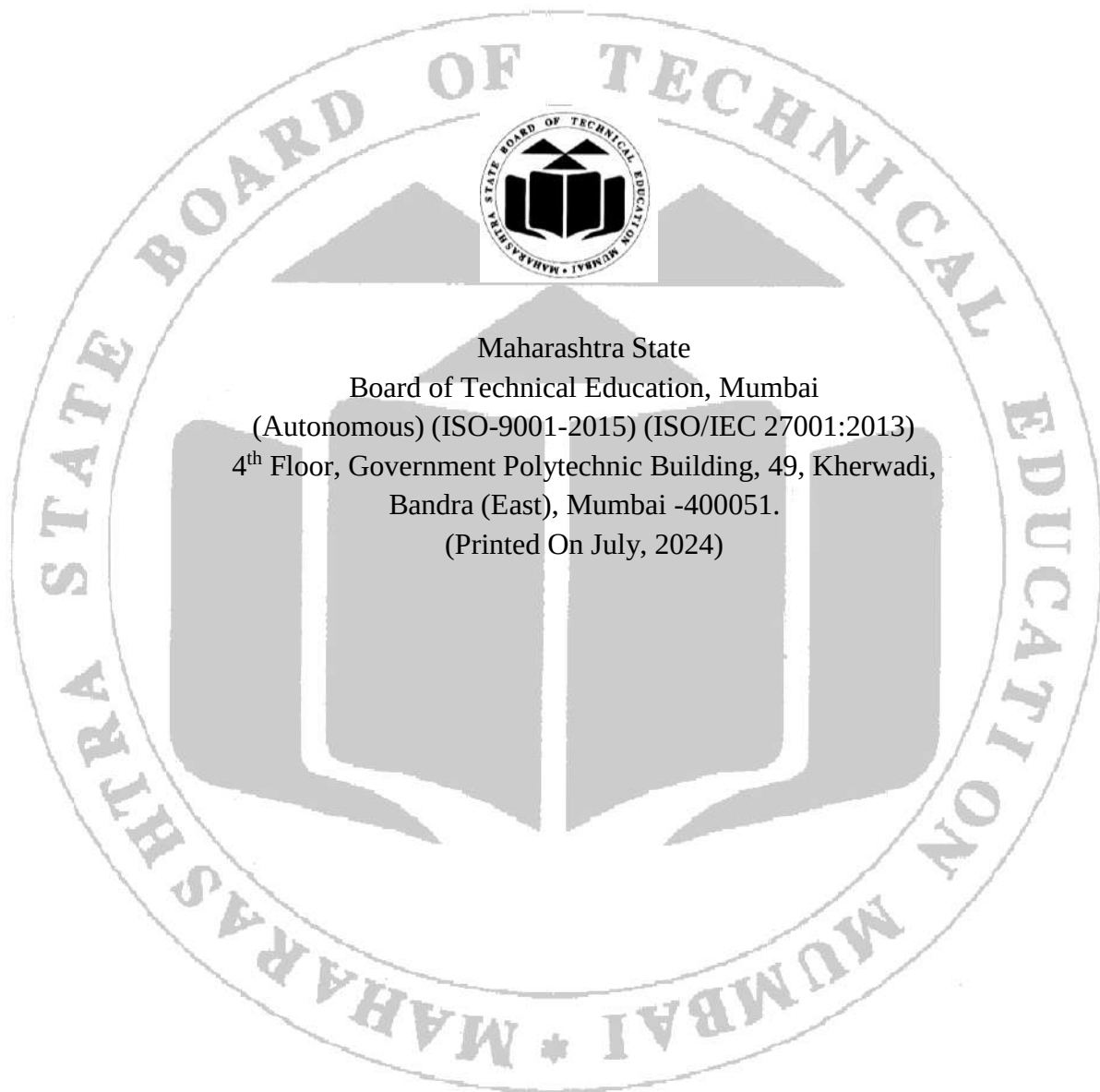
Semester-III

‘K’ Scheme

(AI/AN/DS)



**Maharashtra State
Board of Technical Education, Mumbai**
(Autonomous) (ISO 9001:2015) (ISO/IEC 27001:2013)



Maharashtra State
Board of Technical Education, Mumbai
(Autonomous) (ISO-9001-2015) (ISO/IEC 27001:2013)
4th Floor, Government Polytechnic Building, 49, Kherwadi,
Bandra (East), Mumbai -400051.
(Printed On July, 2024)



**Maharashtra State
Board of Technical Education
Certificate**

This is to certify that Mr. / Ms.....Roll
No., of Third Semester of Diploma in.....
..... of Institute
..... (Institute Code.....) has completed the
Term work satisfactorily in course **Data Story Telling and Visualization
(313004)** for the academic year 20..... To 20.....As prescribed in the
Curriculum.

Place: Enrollment No. :

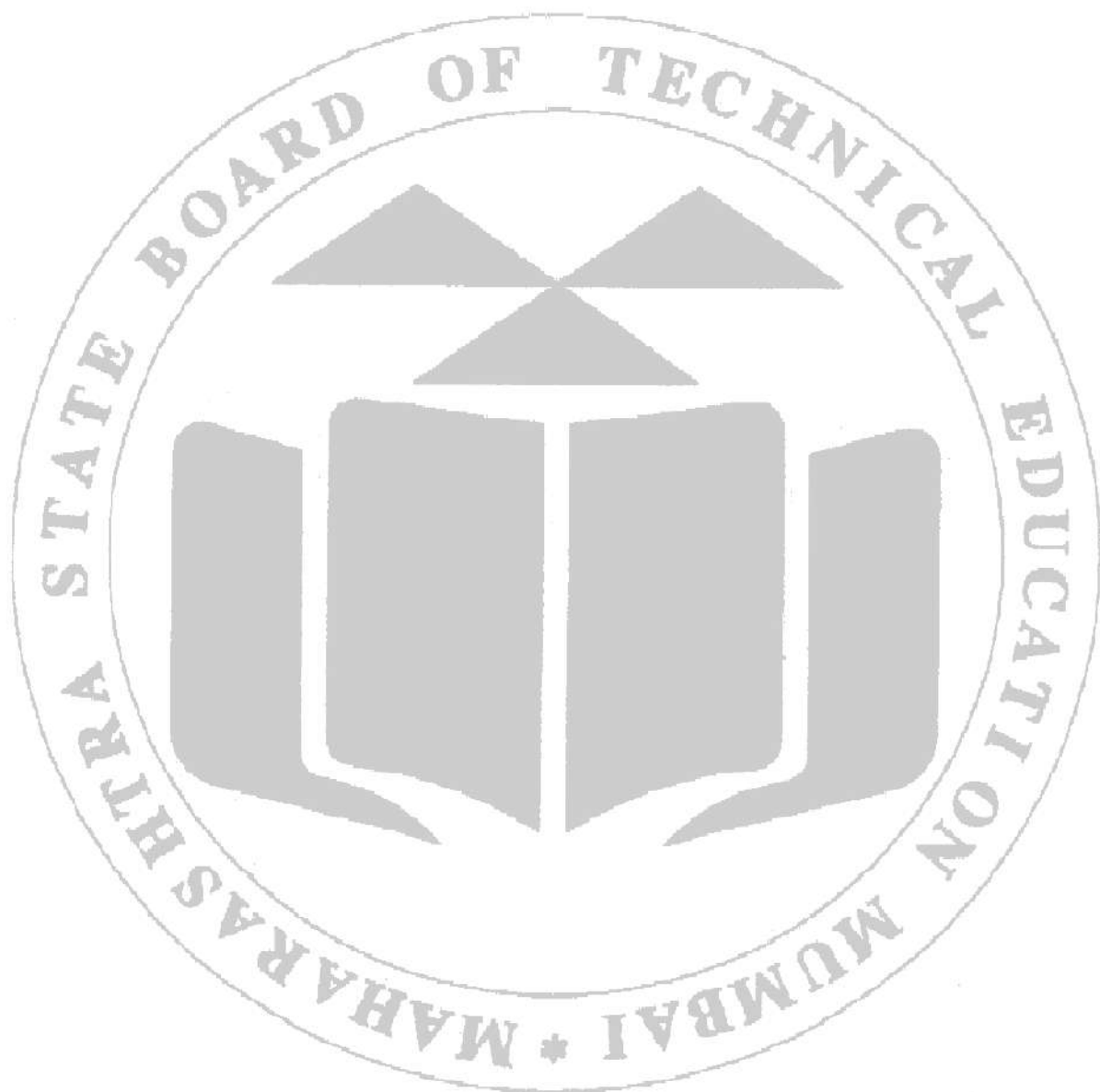
Date: Exam. Seat No:

Subject Teacher

Head of the Department

Principal





Preface

The primary focus of any engineering laboratory/field work in the technical education system is to develop the much needed industry relevant competencies and skills. With this in view, MSBTE embarked on this innovative 'K' Scheme curricula for engineering diploma programs with outcome- based education as the focus and accordingly, relatively large amount of time is allotted for the practical work. This displays the great importance of laboratory work making each teacher, instructor and student to realize that every minute of the laboratory time need to be effectively utilized to develop these outcomes, rather than doing other mundane activities. Therefore, for the successful implementation of this outcome-based curriculum, every practical has been designed to serve as a '*vehicle*' to develop this industry identified competency in every student. The practical skills are difficult to develop through 'Chalk and duster' activity in the classroom situation. Accordingly, the 'K' scheme laboratory manual development team designed the practical to focus on the outcomes, rather than the traditional age old practice of conducting practical to 'verify the theory' (which may become a byproduct along the way).

This laboratory manual is designed to help all stakeholders, especially the students, teachers and instructors to develop in the student the pre-determined outcomes. It is expected from each student that at least a day in advance, they have to thoroughly read through the concerned practical procedure that they will do the next day and understand the minimum theoretical background associated with the practical. Every practical in this manual begins by identifying the competency, industry relevant skills, course outcomes and practical outcomes which serve as a key focal point for doing the practical. The students will then become aware about the skills they will achieve through procedure shown there and necessary precautions to be taken, which will help them to apply in solving real-world problems in their professional life.

This manual also provides guidelines to teachers and instructors to effectively facilitate student- centered lab activities through each practical exercise by arranging and managing necessary resources in order that the students follow the procedures and precautions systematically ensuring the achievement of outcomes in the students.

Learning data Storytelling and visualization is essential for students as it equips them with the skills to effectively communicate complex data insights. By mastering these techniques, students can enhance their ability to analyze and interpret data, present findings persuasively, and influence decision- making processes across various industries. These skills not only broaden career opportunities in fields like data analysis and business intelligence but also foster critical thinking and creativity in data interpretation and presentation.

Although the best possible care has been taken to check for errors (if any) in this laboratory manual, perfection may elude us as this is the first edition of this manual. Any errors and suggestions for improvement are solicited and highly welcome.

Lab Manual Development Team

Program Outcomes (POs) to be achieved through the Practical of this course

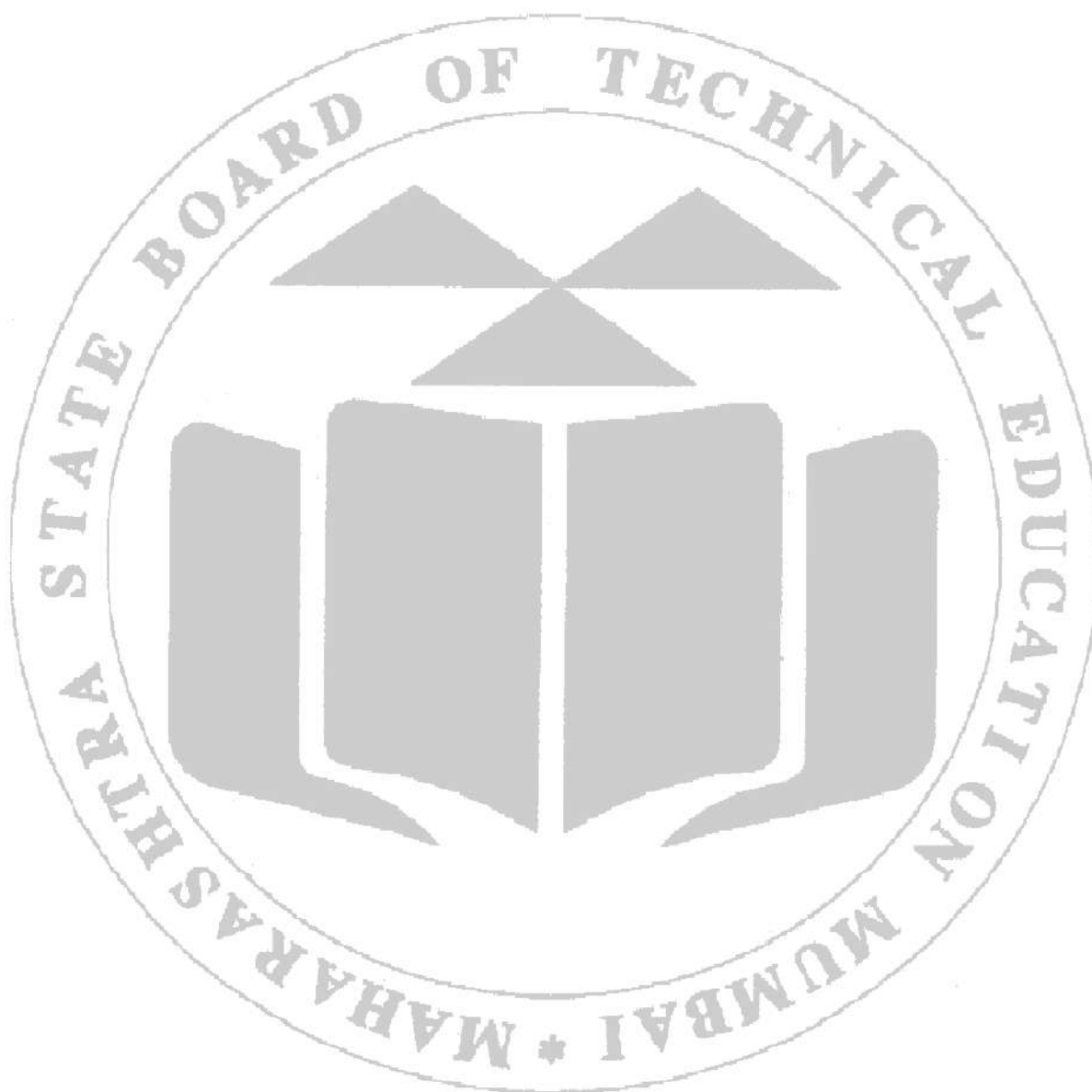
- PO1. Basic and Discipline specific knowledge:** Apply knowledge of basic mathematics, science and engineering fundamentals and engineering specialization to solve the engineering problems.
- PO2. Problem analysis:** Identify and analyze well-defined engineering problems using codified standard methods.
- PO3. Design/ development of solutions:** Design solutions for well-defined technical problems and assist with the design of systems components or processes to meet specified needs.
- PO4. Engineering Tools, Experimentation, and Testing:** Apply modern engineering tools and appropriate techniques to conduct standard tests and measurements.
- PO5. Engineering practices for society, sustainability, and environment:** Apply appropriate technology in the context of society, sustainability, environment, and ethical practices.
- PO6. Project Management:** Use engineering management principles individually, as a team member, or as a leader to manage projects and effectively communicate about well-defined engineering activities.
- PO7. Life-long learning:** Ability to analyze individual needs and engage in updating in the context of technological changes.

Practical- Course Outcome matrix

Course Outcomes (COs) CO1. Identify the characters in data Storytelling. CO2. Eliminate clutter and grab audience attention. CO3. Construct Storytelling for the given incident. CO4. Transform Data to Visuals. CO5. Create data visualization using many distributions.						
Sr. No.	Title of the Practical	CO1	CO2	CO3	CO4	CO5
1	Identify a project you are working on where you need to communicate in a data-driven way. Reflect upon and write the answers to the following questions. i) Who is your Audience? ii) What does your audience care about? iii) What action does your audience need to take?	√				
2	* Identify a project you are working on where you need to communicate in a data-driven way. Reflect upon and write the answers of following questions. a) What is a Stake? i) What are the benefits if your audience acts in the way that you want them to? ii) What are the risks if they do not? b) Form a Big Idea i) Articulate your point of view. ii) Convey what's at stake. iii) Be a complete (and single) sentence.	√				
3	* Determine audience, communication mechanism, desired tone and Select Effective Visuals for any incident (<i>Below is an example for reference only</i>). <i>Teacher shall allocate similar assignments.</i> i. Who is your Audience? ii. List the primary groups or individuals to whom you'll be communicating. iii. If you had to narrow that to a single person, who would that be? iv. What does your audience care about? v. What action does your audience need to take? <i>Example: Buses Bunching - Make a data story on bus bunching, (Bus Bunching means when a bus gets delayed and later causes multiple buses to arrive at a single stop at the same time.</i>	√				

4	* Make a clutter free Data Story on any incident. - Identify and eliminate clutter. - Select suitable pre-attentive attributes. - Explore affordances, accessibility, and Aesthetics as per designers view.		√	√		
5	Create a simple video (up to 3 minutes) telling a story on incidence given in Practical no.3	√	√	√		
6	* Create a data story for Vehicle (Bicycle/Bike/Car/Bus etc.) Rental System. (Below is an example for reference only). Teacher shall allocate similar assignments. Example: Create a data story with respect to the following observations: - What are the most popular pick-up locations across the city for Citi Bike rental? - How does the average trip duration vary across different age groups, and over time? - Which age group rents the most bikes? - How does bike rental vary across the two user groups (one-time users vs. long-term subscribers) on different days of the week? - Do factors like weather and user age impact the average bike trip duration?	√	√	√		
7	* Create a video (up to 5 minutes) telling a story on given Incidence. Record a video of yourself speaking, or narrate while showing visual props or sketches, or screencast a PowerPoint presentation, etc. Choose how you present the story. Produce a single video file (formatted as a .mov file) Teacher shall suggest various incidents to the students	√	√	√		
8	Create a bar chart for data visualizations on Practical No. 6.	√	√	√		
9	* Construct a Case study on data Story telling for any Musical/Social App. Example: Spotify takes the data from our listening habits and spins it into an exciting audio and visual experience. Teacher shall allocate similar case study.		√	√	√	
10	* Implementation of a python program that performs data cleaning on any dataset.		√	√	√	
11	Create Bar chart for data visualization using Single distribution. Example: Histogram of the ages of the train/flight passengers. Teacher shall allocate similar assignments.				√	√
12	Develop a worksheet, add filters and create chart using dataset by using any Visualization tool.				√	√

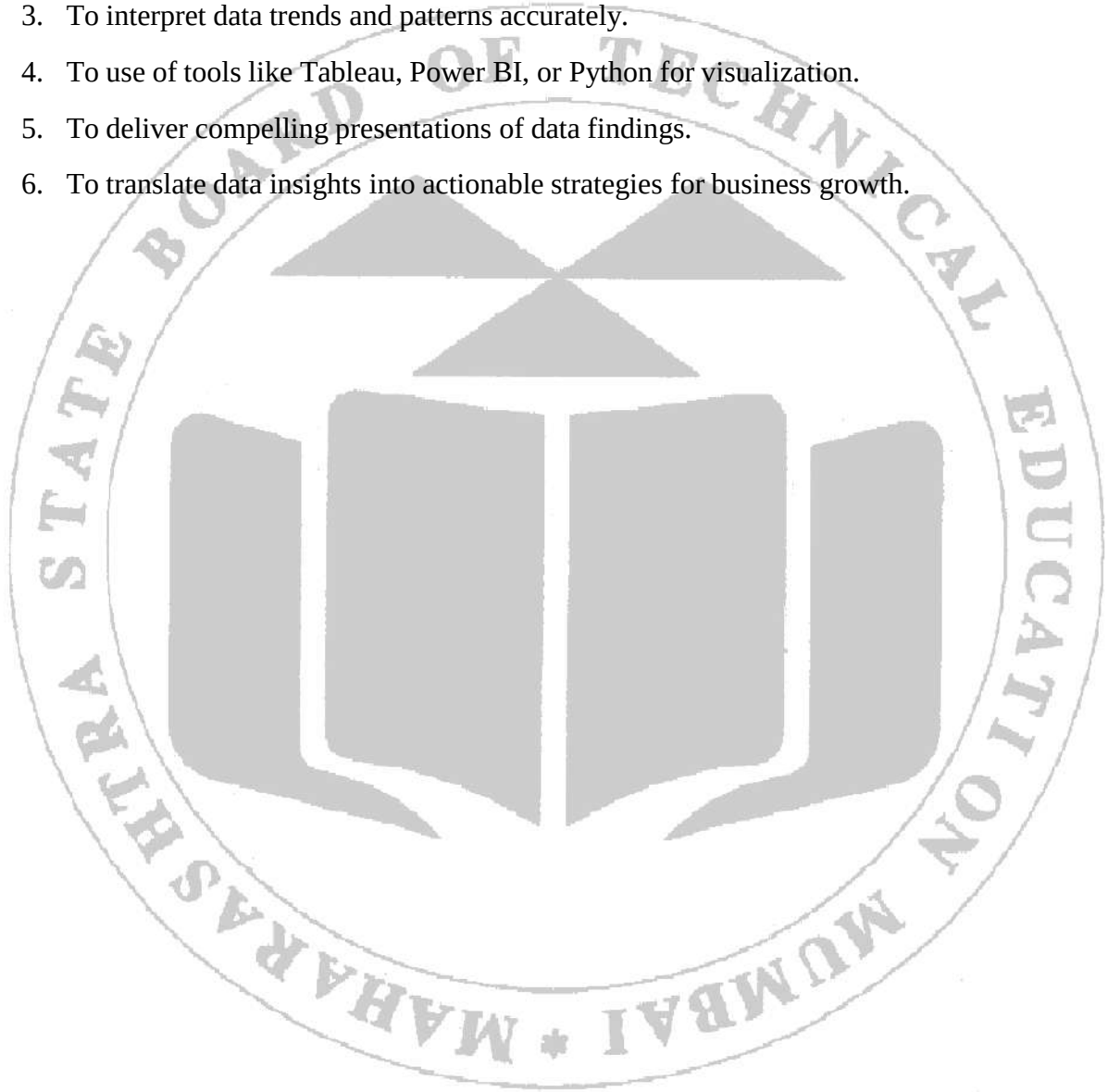
13	* Create Bar chart for data visualization using Many distribution. <i>Example: Histogram of the gender and ages of the train/flight passengers. Teacher shall allocate similar assignments.</i>				√	√
14	Implementation of a python program that loads a dataset and plot grouped bars.			√	√	√
15	Implementation of a python program that loads any dataset and plot a pie chart.					√



Industry / Employer Expected Outcome

The aim of this course is to help the student to attain the following industry identified outcomes through various teaching learning experiences:

1. To extract the valuable insights from complex data sets
2. To develop proficiency in creating impactful charts, graphs, and dashboards.
3. To interpret data trends and patterns accurately.
4. To use of tools like Tableau, Power BI, or Python for visualization.
5. To deliver compelling presentations of data findings.
6. To translate data insights into actionable strategies for business growth.



Guidelines to Teachers

1. For incidental writing on the day of each practical session every student should maintain a dated logbook for the whole semester, apart from this laboratory manual, which s/he has to submit for assessment to the teacher in the next practical session.
2. Teachers should give opportunity to students for hands-on after the demonstration.
3. Assess the skill achievement of the students and COs of each unit.
4. Explain prior concepts to the students before starting of each experiment.
5. List of few sample questions for reference are given. Teachers must design more such questions so as to ensure the achievement of identified CO.
6. Teacher should ensure that the practical skill and competencies are developed in the students after the completion of the practical exercise.
7. Teacher may provide additional knowledge and skills to the students even though it's not covered in the manual but are expected from the students by the industries.
8. Teacher may suggest the students to refer additional related literature of the Technical papers/ Reference books/ Seminar proceedings, etc.
9. Teacher shall assess the performance of students continuously as per norms prescribed by MSBTE.
10. During assessment teacher is expected to ask questions to the students to tap their Achievements grading related knowledge and skills. So that, student can prepare while submitting record of the practical focus should be given on development of enlisted skills rather than theoretical knowledge.

Instructions for Students

1. Understand the purpose of practical and its implementation.
2. Student shall develop practical skills as expected by the Industries.
3. Listen carefully to the instructions given by the teacher about importance of relevant program Outcomes, relevant course outcomes, practical significance, competency and practical skills, practical outcome and the theoretical background during the practical session.
4. Write the answers of the questions allotted by the teacher during practical session.
5. Student should develop the habit of group discussion related to the practical, so that exchange of knowledge/skills could take place.
6. Student shall attempt to develop related hands-on-skills to gain confidence.
7. Student shall refer technical magazines, websites related to the scope of the course.
8. Student should develop habit to submit the practical, exercise continuously and progressively on the scheduled dates and should get the assessment done.
9. Student should be well prepared while submitting the write up of the exercise.
10. Student should not hesitate to ask any difficulty faced during conduct of practical.

Content Page**List of Practical and Progressive Assessment Sheet**

Sr. No.	Title of the Practical	Page No	Date of Performance	Date of submission	Assessment marks (25)	Dated sign. of teacher	Remarks (if any)
1	Identify a project you are working on where you need to communicate in a data-driven way. Reflect upon and write the answers to the following questions. i. Who is your Audience? ii. What does your audience care about? iii. What action does your audience need to take?	1					
2	* Identify a project you are working on where you need to communicate in a data-driven way. Reflect upon and write the answers of following questions. a) What is a Stake? i. What are the benefits if your audience acts in the way that you want them to? ii. What are the risks if they do not? b) Form a Big Idea i. Articulate your point of view. ii. Convey what's at stake. iii. Be a complete (and single) sentence.	6					
3	Determine audience, communication mechanism, desired tone and Select Effective Visuals for any incident (<i>Below is an example for reference only</i>). Teacher shall allocate similar assignments. i. Who is your Audience? ii. List the primary groups or individuals to whom you'll be communicating. iii. If you had to narrow that to a single person, who would that be? iv. What does your audience care about? v. What action does your audience need to take? <i>Example: Buses Bunching - Make a data story on bus bunching, (Bus Bunching means when a bus gets delayed and later causes multiple buses to arrive at a single stop at the same time.</i>	10					

4	* Make a clutter free Data Story on any incident. - Identify and eliminate clutter. - Select suitable pre-attentive attributes. - Explore affordances, accessibility, and aesthetics as per designers view.	14					
5	Create a simple video (up to 3 minutes) telling a story on incidence given in Practical no.3	19					
6	* Create a data story for Vehicle (Bicycle/Bike/Car/Bus etc.) Rental System. (Below is an example for reference only). Teacher shall allocate similar assignments. <i>Example: Create a data story with respect to the following observations:</i> - What are the most popular pick-up locations across the city for Citi Bike rental? - How does the average trip duration vary across different age groups, and over time? - Which age group rents the most bikes? - How does bike rental vary across the two user groups (one-time users vs. long-term subscribers) on different days of the week? - Do factors like weather and user age impact the average bike trip duration?	24					
7	* Create a video (up to 5 minutes) telling a story on given Incidence. Record a video of yourself speaking, or narrate while showing visual props or sketches, or screencast a PowerPoint presentation, etc. Choose how you present the story. Produce a single video file (formatted as a .mov file) Teacher shall suggest various incidents to the students.	29					
8	Create a bar chart for data visualizations on Practical No. 6.	33					
9	* Construct a Case study on data Story telling for any Musical/Social App. <i>Example: Spotify takes the data from our listening habits and spins it into an exciting audio and visual experience.</i> <i>Teacher shall allocate similar case study.</i>	38					
10	* Implementation of a python program that performs data cleaning on any dataset.	45					

11	Create Bar chart for data visualization using Single distribution. <i>Example: Histogram of the ages of the train/flight passengers. Teacher shall allocate similar assignments.</i>	49					
12	Develop a worksheet, add filters and create chart using dataset by using any Visualization tool.	54					
13	* Create Bar chart for data visualization using Many distribution. <i>Example: Histogram of the gender and ages of the train/flight passengers. Teacher shall allocate similar assignments.</i>	59					
14	Implementation of a python program that loads a dataset and plot grouped Bars.	65					
15	Implementation of a python program that loads any dataset and plot a pie chart.	70					
Total Marks							
(Note : Out of above suggestive LLOs - • '*' Marked Practical (LLOs) Are mandatory. • Minimum 80% of above list of lab experiment are to be performed. • Judicial mix of LLOs are to be performed to achieve desired outcomes.							

Note: Marks to be transferred to Performa of (CIAAN-K format).

Practical No.1: Identify a project you are working on where you need to communicate in a data-driven way (Part-1).

I Practical Significance

By learning to tell data stories effectively, students can not only enhance their analytical skills but also improve their ability to communicate insights in a way that drives real-world actions and outcomes. Data Story telling is crucial for improving employee productivity by translating complex data into clear, actionable insights. For senior management, it enables informed strategic decisions and effective resource allocation, directly impacting business performance. The HR department benefits from targeted programs and proactive problem-solving by understanding productivity trends and employee satisfaction. Team leaders and department heads gain operational efficiency, better performance management, and enhanced collaboration. Overall, data Storytelling fosters a unified understanding, making insights actionable and supporting a culture of continuous improvement across the organization.

II Industry/ Employer Expected Outcome

1. To extract the valuable insights from complex data sets
2. To develop proficiency in creating impactful charts, graphs, and dashboards

III Course Level Learning Outcome

CO1-Identify the characters data storytelling.

IV Laboratory Learning Outcome

LLO 1.1-Identify the audience.

LLO1.2-Interpret the needs of the audience

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

By presenting data in a clear and engaging manner, it ensures that each audience group can understand, appreciate, and act upon the insights, leading to sustained organizational success. You should be Specific of who your audience is for effective Communication. Narrow your Target Audience by identifying the decision maker. Avoid General Audience. You can satisfy the needs of your Audience then you know more about your Audience. Audience should trust you as an expert.

VII Exercises

Follow the procedure given below:

1. **Project Choice:** Pick a topic for your project where data can help to tell a story.

2. **Understand Audience:** Know who you want to share the story with.
3. **Audience Needs:** Find out what information your audience wants.
4. **Story Design:** Create a clear story that fits what your audience needs.
5. **Visualize Data:** Use charts or pictures to explain the data.
6. **Feedback:** Show your story to people and get their thoughts to improve.

VIII Procedure with Example Scenario for Students: Analyzing Exam Results to Improve Academic Performance

Imagine a student working on a project to analyze exam results in a hypothetical high school. They gather data on various factors such as study habits, attendance records, extracurricular involvement, and previous academic performance. Using data Storytelling, the student can:

Create a Narrative: Develop a story that shows a correlation between consistent study habits and higher exam scores. The narrative can highlight how students who maintain regular study schedules tend to perform better academically.

Use Visual Aids: Utilize visual aids like scatter plots, bar graphs, and pie charts to illustrate the impact of different factors on exam performance. For example, a scatter plot can show the relationship between hours spent studying and exam scores, while a pie chart can depict the distribution of scores across different attendance levels.

Present Recommendations: Based on the data analysis, the student can recommend strategies such as implementing study skill workshops, offering after-school tutoring programs, or encouraging regular attendance through incentive programs.

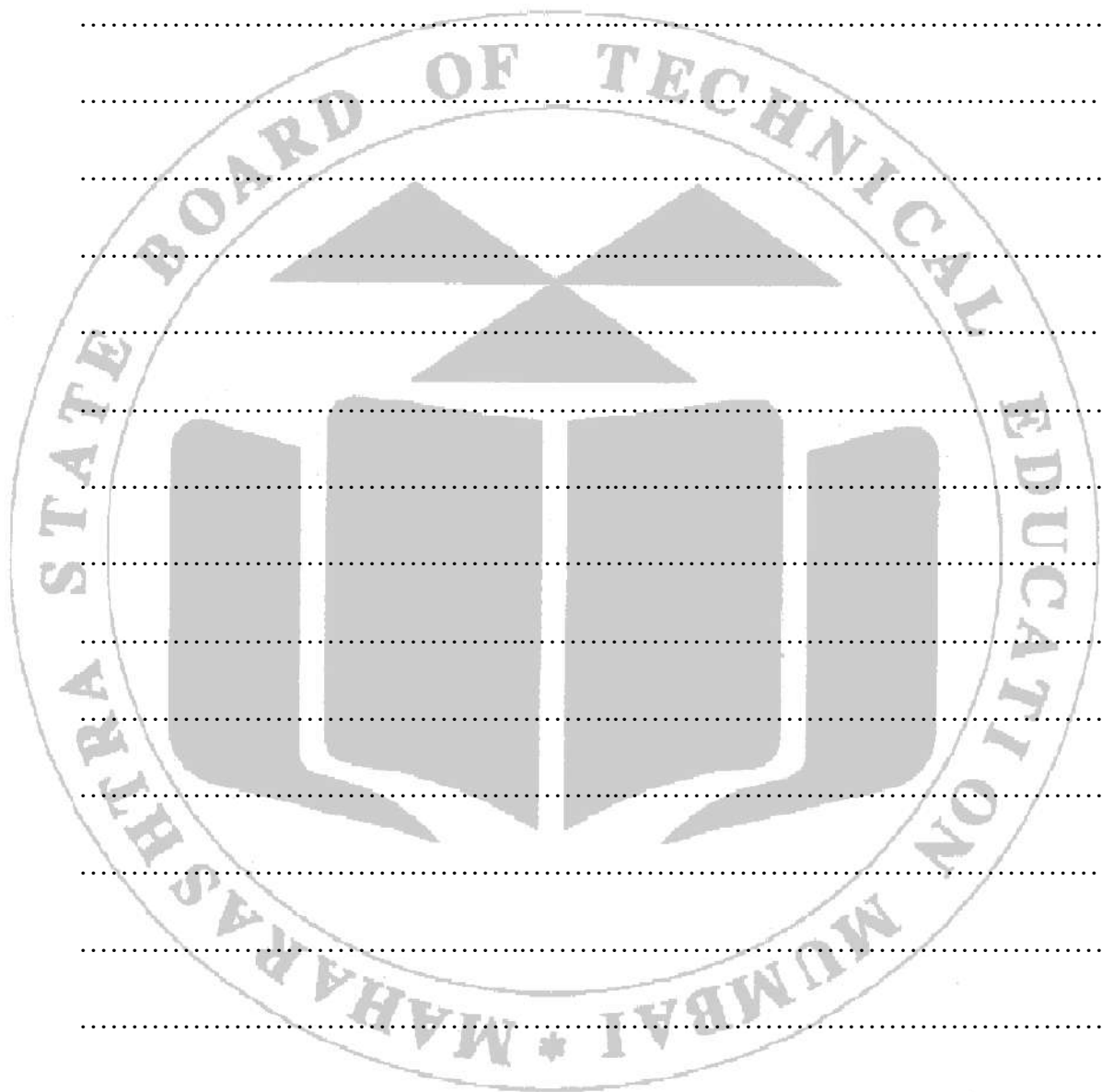
Highlight ROI: Demonstrate the potential return on investment (ROI) of these recommendations. For instance, the student can calculate the expected improvement in overall school performance and the long-term benefits of better academic outcomes, such as higher college acceptance rates and scholarships.

IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be followed

1. **Data Integrity:** Ensure the accuracy and reliability of the data used for visualization to avoid misleading insights or conclusions.
2. **Ethical Representation:** Present data and visualizations ethically, avoiding distortion or manipulation that could lead to misinterpretation or biased narratives.



XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.datapine.com/blog/data-Storytelling-examples-with-data-visualization/>
2. <https://www.youtube.com/watch?v=ZFGG2ppSlao>
3. <https://powerbi.microsoft.com/en-in/data-Story telling/>
4. <https://www.youtube.com/watch?v=vD6CdzM8ZZw>

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.2: Identify a project you are working on where you need to communicate in a data-driven way (Part-2).

I Practical Significance

The practical significance of this exercise lies in its ability to enhance your communication and persuasion skills in data-driven projects. By clearly identifying what is at stake, you can articulate the benefits and risks in a way that compels your audience to act in your favor. This clarity helps in crafting a concise and memorable big idea that aligns complex information with strategic objectives. Moreover, understanding potential risks allows for better risk management and contingency planning, ensuring project resilience. A well-formed big idea ensures focused implementation and alignment among team members, driving efficient project execution. Ultimately, this exercise fosters accountability and sets measurable expectations for project outcomes, significantly contributing to the overall success and impact of your project.

II Industry/ Employer Expected Outcome

1. To extract the valuable insights from complex data sets
2. To develop proficiency in creating impactful charts, graphs, and dashboards.

III Course Level Learning Outcome

CO1-Identify the characters data storytelling.

IV Laboratory Learning Outcome

LLO 2.1-Identify the stake.

LLO 2.2-Design the big idea

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

This exercise emphasizes the importance of effective communication in data-driven projects. Clear expression of what is at stake, including benefits and risks, enhances persuasion and influence over stakeholders. A well-formed big idea distills complex information into a concise, memorable message that aligns with strategic objectives. This clarity aids in focused implementation and risk management, ensuring project resilience. Effective communication fosters alignment and accountability among team members, driving efficient project execution.

VII Exercises

Follow the procedure given below:

1. **Project Choice:** Pick a topic for your project where data can help to tell a story.
2. **Understand Audience:** Know who you want to share the story with.
3. **Audience Needs:** Find out what information your audience wants.

4. **Story Design:** Create a clear story that fits what your audience needs.
5. **Visualize Data:** Use charts or pictures to explain the data.
6. **Feedback:** Show your story to people and get their thoughts to improve.

VIII Procedure with Example: New Customer Analytics Dashboard

- **Stake:**
 1. **Benefit if they act:** The sales team will increase sales by 20% using real-time customer data for targeted offers.
 2. **Risk if they don't:** Sales could drop by 10% due to untargeted marketing without this tool.
- **Big Idea:**
 1. **Point of View:** The new dashboard will enable precise, data-driven marketing decisions.
 2. **What's at Stake:** Adopting it will enhance campaign success and customer engagement, while ignoring it risks losing touch with customer needs.
 3. **Complete Sentence:** Using the new customer analytics dashboard will enhance our marketing effectiveness and customer engagement, while not using it could lead to missed insights and lower engagement.

IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be followed

1. **Data Integrity:** Ensure the accuracy and reliability of the data used for visualization to avoid misleading insights or conclusions.
2. **Ethical Representation:** Present data and visualizations ethically, avoiding distortion or manipulation that could lead to misinterpretation or biased narratives.

XI Conclusion

.....

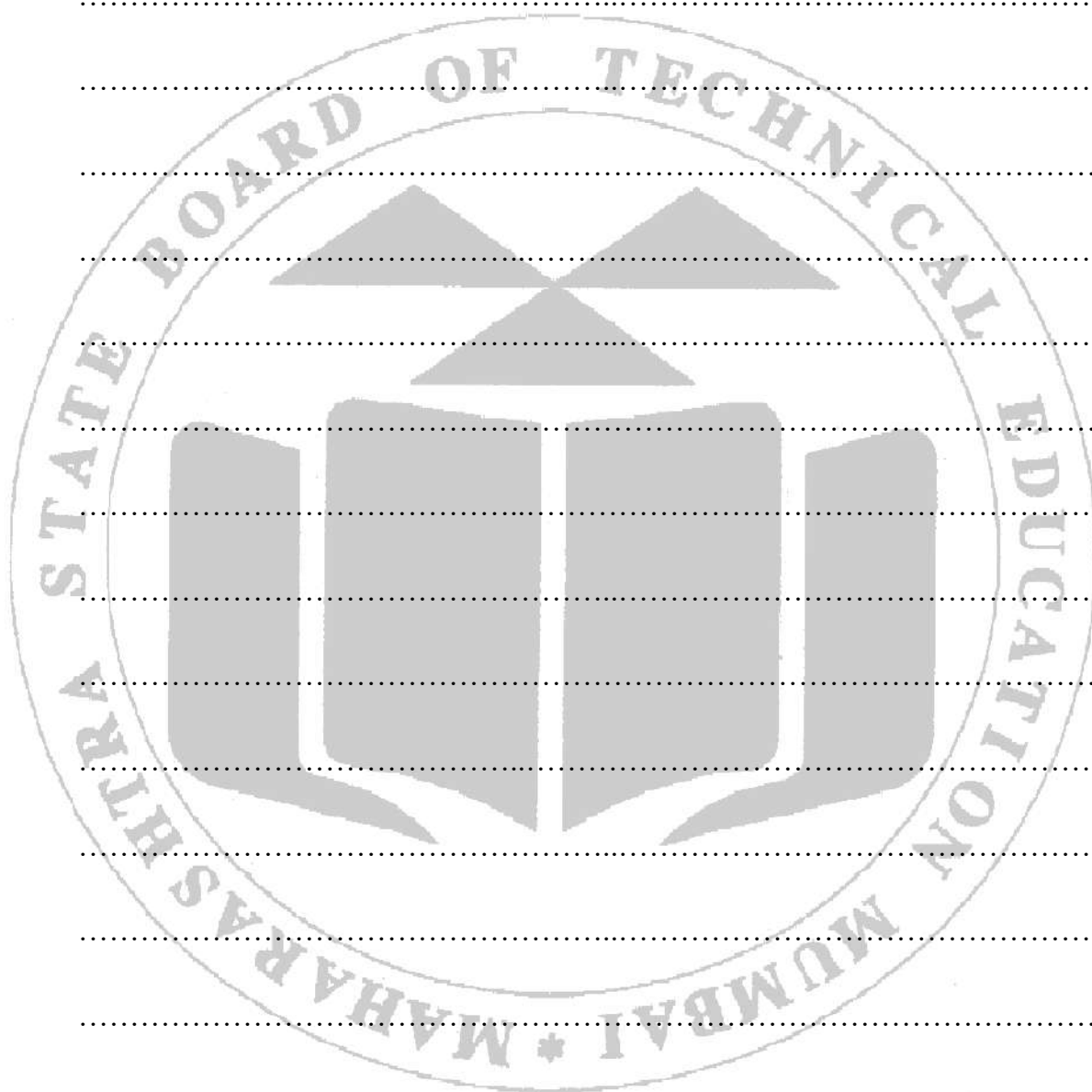
.....

XII Practical Related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

- a) What is a Stake?
 - i) What are the benefits if your audience acts in the way that you want them to?
 - ii) What are the risks if they do not?

[Space for Answer]



XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.datapine.com/blog/data-Storytelling-examples-with-data-visualization/>
2. <https://powerbi.microsoft.com/en-in/data-Storytelling/>
3. <https://www.youtube.com/watch?v=vD6CdzM8ZZw>

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.3: Determine audience, communication mechanism, desired tone and Select Effective Visuals for any incident (e.g. bus bunching)

I Practical Significance

The practical significance of effectively communicating about bus bunching lies in its potential to enhance public transportation efficiency and traveler satisfaction. By engaging with diverse stakeholders—from daily travelers seeking reliable service to city officials responsible for policy and funding decisions—clear communication can foster understanding of the issue's impact and urgency. Using targeted visuals and a tailored tone ensures that each group comprehends the implications and is motivated to take actions, whether through operational improvements, policy adjustments, or public awareness initiatives, thereby contributing to a more efficient and responsive public transportation system overall.

II Industry/ Employer Expected Outcome

1. To translate data insights into actionable strategies for business growth.
2. To develop proficiency in creating impactful charts, graphs, and dashboards

III Course Level Learning Outcome

CO1-Identify the characters data storytelling.

IV Laboratory Learning Outcome

LLO 3.1-Construct data story on a given incident

LLO 3.2-Determine the audience, communication mechanism, and desired tone.

LLO 3.3-Decide effective visuals.

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

Bus bunching occurs when delays in one bus cause multiple buses to arrive simultaneously at a stop. This phenomenon disrupts schedules, leading to longer wait times and overcrowded buses, frustrating commuters and impacting overall service reliability. Understanding and addressing bus bunching through data-driven analysis and effective communication is crucial for improving public transportation efficiency and passenger satisfaction.

VII Exercises

Write the data Story on Bike Sharing Availability

VIII Procedure with Example: Data Story on Bus Bunching

Bus bunching occurs when delays in one bus cause subsequent buses to arrive

simultaneously at a single stop, disrupting schedules and inconveniencing commuters. This phenomenon is not only frustrating for passengers but also indicates inefficiencies within the public transportation system.

Audience:

- **Primary Audience:** Daily commuters reliant on bus services, transit authorities, and city transportation officials.
- **Single Individual:** The head of the city's transportation department.

Key Points:

- **Impact on Commuters:** Visualizations show how bus bunching affects travel times and passenger waiting experiences.
- **Causes and Frequency:** Data graphs illustrate common causes of delays (e.g., traffic congestion, mechanical issues) and frequency of bus bunching incidents.
- **Operational Efficiency:** Comparison charts between scheduled and actual bus arrivals highlight inefficiencies in service operations.
- **Mitigation Strategies:** Recommendations for improving bus scheduling, operational adjustments, and real-time monitoring solutions.

Desired Actions:

- **Commute Users:** Awareness of bus bunching impacts and encouragement to provide feedback on service experiences.
- **Transportation Authorities:** Implementation of strategies to reduce bunching through improved scheduling and operational protocols.
- **City Officials:** Support for funding and policy changes aimed at enhancing public transit reliability and efficiency.

Visuals:

- Info graphics illustrating delays and bunching incidents.
- Comparative charts showing scheduled vs. actual bus arrivals.
- Maps depicting routes and locations affected by bunching.
- Graphs displaying trends and frequency of bunching over time.

By presenting these insights effectively, this data story aims to inform stakeholders, shoot actionable insights, and ultimately contribute to a more reliable and efficient public transportation system.

IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

XIII References / Suggestions for further Reading Software/Learning Websites

1. [https://www.datapine.com/blog/data-Story telling-examples-with-data-visualization/](https://www.datapine.com/blog/data-Story-telling-examples-with-data-visualization/)
2. <https://setosa.io/bus/>
3. https://en.wikipedia.org/wiki/Bus_bunching
4. https://www.youtube.com/watch?v=6X_YsPketmk

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.4: Make a clutter free Data Story on any incident.

- i. Identify and eliminate clutter.**
- ii. Select suitable pre-attentive attributes.**
- iii. Explore affordances, accessibility, and aesthetics as per designers view.**

I Practical Significance

Creating a clutter-free data story holds significant practical importance across various domains such as business, academia, journalism, and public policy. It enhances communication by presenting information clearly and efficiently, making key insights more accessible and understandable. This clarity supports informed decision-making, increases viewer engagement, and encourages exploration of the data without overwhelming the audience. Additionally, a clutter-free approach improves accessibility for diverse audiences and enhances the professional presentation of information, ultimately boosting credibility and trust in the data presented.

II Industry/ Employer Expected Outcome

1. To translate data insights into actionable strategies for business growth.
2. To develop proficiency in creating impactful charts, graphs, and dashboards

III Course Level Learning Outcome

- CO2-Eliminate clutter and grab audience attention.
CO3-Construct Story telling for the given incident.

IV Laboratory Learning Outcome

- LLO 4.1- Construct a data story on a given incident.
LLO 4.2-Perform de-cluttering.
LLO 4.3-Choose suitable pre-attentive attributes.
LLO 4.4-Determine feasibility as per the designer's view

V Relevant Affective domain related Outcome(s)

3. Follow safety practices.
4. Maintain tools and equipment.
5. Follow ethical practices.

VI Relevant Theoretical Background

Clutter in the context of data visualization or Story telling refers to unnecessary or excessive elements that hinder understanding or distract from the main message. Clutter can take various forms: *visual clutter*, *textual clutter*, *data clutter*. Thus, prioritizing simplicity and clarity in data Story telling not only improves comprehension but also facilitates effective communication and decision-making processes across different sectors.

VII Exercises

Follow the procedure given below:

- **Define the Problem:** Clearly articulate the issue of bus bunching.
- **Data Collection and Preparation:** Obtain historical or real-time data on bus arrivals, departures, and locations.
- **Data Analysis:** Conduct time-series analysis to identify patterns and irregularities in bus arrivals.
- **Spatial Analysis and Mapping:** Identify specific locations prone to bunching and possible reasons for it.

VIII Procedure with Example:

Incident: COVID-19 Vaccination Progress in a Country.

Creating a clutter-free data story involves several key steps to ensure clarity and effectiveness in communication. Let's outline these steps using a hypothetical incident for illustration:

i) Identify and Eliminate Clutter:

1. **Focus on Key Metrics:** Select the most relevant metrics to convey the vaccination progress, such as total doses administered, percentage of population vaccinated, and daily vaccination rates.
2. **Simplify Visual Elements:** Use minimalistic design principles. Avoid unnecessary decorations, backgrounds, or excessive colors that can distract from the data.
3. **Reduce Text:** Use concise labels and captions. Eliminate jargon or unnecessary explanations that can clutter the visual presentation.
4. **Remove Non-Essential Data:** Exclude data points or graphs that do not directly contribute to understanding the vaccination progress.

ii) Select Suitable Pre-Attentive Attributes:

1. **Color:** Use color strategically to highlight important data points (e.g., use green for vaccinated population, red for active cases).
2. **Size:** Represent data significance through size (e.g., larger circles for higher vaccination rates).
3. **Position:** Arrange elements logically to guide the viewer's attention (e.g., chronological order of vaccination milestones).

iii) Explore Affordances, Accessibility, and Aesthetics:

1. **Affordances:** Ensure interactive elements like hover effects for detailed information and clickable links for further resources.
2. **Accessibility:** Use high contrast between text and background for readability. Provide alternative text for images and graphs for screen readers.
3. **Aesthetics:** Balance visual appeal with functionality. Choose a clean layout with ample white space to avoid visual overload.

Example Visualization:

- **Title:** COVID-19 Vaccination Progress in Country X
- **Visualization Type:** Interactive line graph showing daily vaccination rates over time, with overlays for key milestones (e.g., 50% vaccinated).
- **Color Scheme:** Green for vaccinated, gray for total population, with a subtle gradient background.
- **Interactivity:** Hover over points for exact figures, click for detailed weekly summaries.

IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be Followed

1. Data Integrity: Ensure the accuracy and reliability of the data used for visualization to avoid misleading insights or conclusions.
2. Ethical Representation: Present data and visualizations ethically, avoiding distortion or manipulation that could lead to misinterpretation or biased narratives.

XI Conclusion

.....

.....

XII Practical Related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

- What is Clutter?
- How to identify and eliminate the clutter?

[Space for Answer]

.....

.....

.....

.....

.....

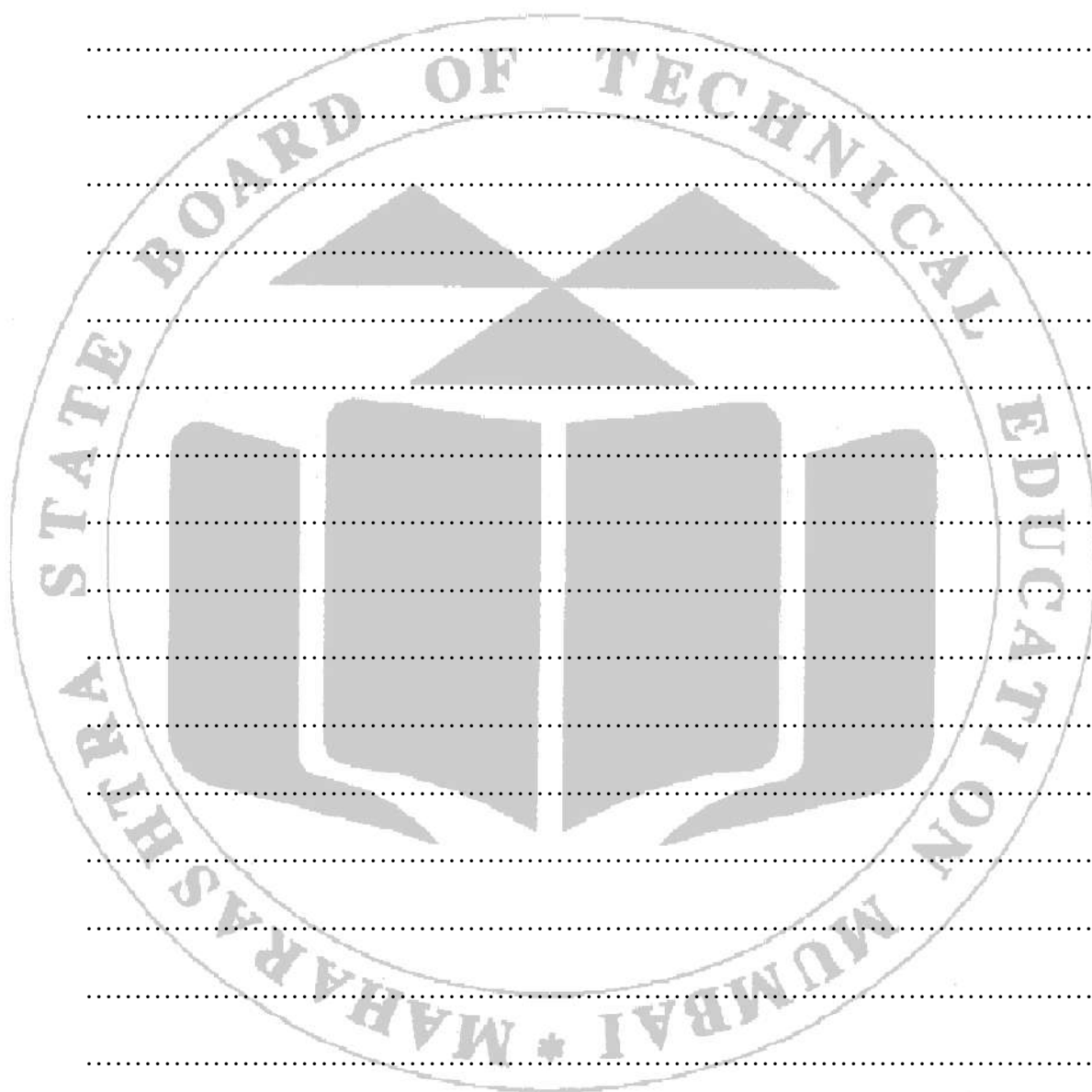
.....

.....

.....

.....

.....



XIII References / Suggestions for further Reading Software/Learning Websites

1. https://www.youtube.com/watch?v=jOfIcKkKWSQ&ab_channel=DatasaurusRex
2. https://www.youtube.com/watch?v=XgY1-BsZSzM&ab_channel=Storytellingwithdata
3. https://www.youtube.com/watch?v=4mka0AZUVMk&ab_channel=RADACAD

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.5: Create a simple video (up to 3 minutes) telling a story on incidence given in Practical no.3 (e.g. bus bunching.)

I Practical Significance

By working through this practical, students gain hands-on experience in Storytelling and visualization techniques to extract meaningful insights from raw data. Creating a simple video about bus bunching is important because it raises awareness, educates the public, and engages a wide audience through visual Storytelling. By explaining the causes and impacts of bus bunching, the video can help passengers understand the issue and encourage transit authorities and policymakers to implement solutions. Additionally, such a video can spark public discourse, promoting collective efforts to improve public transportation systems, ultimately leading to more reliable and efficient transit services for everyone.

II Industry/ Employer Expected Outcome

1. To translate data insights into actionable strategies for business growth.
2. To extract the valuable insights from complex data sets
3. To develop proficiency in creating impactful charts, graphs, and dashboards

III Course Level Learning Outcome

- CO1-Identify the characters in data Storytelling.
- CO2-Eliminate clutter and grab audience attention.
- CO3-Construct Story telling for the given incident.

IV Laboratory Learning Outcome

- LLO 5.1-Compose a video for data Story telling

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

Creating a video up to 3 minutes long to tell a practical story involves structuring a compelling narrative that engages viewers through visual and auditory elements. It's about organizing content cohesively, using visuals effectively to clarify and evoke emotions, and understanding the audience to tailor the message. Practical application is key, ensuring the story's relevance and encouraging viewers to apply its insights. Concluding with a strong message or call to action prompts reflection and potential action, aiming to leave a lasting impact on the audience. Here are some steps to follow while creating the video:

- i. **Data Selection:** Choose a specific dataset or topic for analysis.
- ii. **Storyline Design:** Develop a clear narrative structure, outlining key points.
- iii. **Visualization Creation:** Create visual representations (charts, graphs) to illustrate data insights.
- iv. **Script Writing:** Write a concise and engaging script, aligning narration with visuals.
- v. **Narration Recording:** Record a clear and articulate narration of the script.

- vi. **Video Compilation:** Combine visuals and narration using video editing software.
- vii. **Editing:** Ensure smooth transitions, visual clarity, and proper pacing.
- viii. **Review and Refinement:** Fine-tune narration, visuals, and overall coherence.
- ix. **Finalizing:** Add intro/outro, subtitles, and citations as needed.
- x. **Testing:** Ensure video is easily understandable and visually appealing.
- xi. **Publishing:** Share the video on relevant platforms for intended audience.

VII Exercises

Create a simple video (up to 3 minutes) telling a story on bus bunching.

VIII Procedure with Example:

Creating a video involves several steps, including scripting, storyboarding, recording, and editing. Here's a simple script and storyboard outline for a 3-minute video on the concept of bus bunching.

Script

Scene 1: Introduction (0:00 - 0:30)

- **Visual:** Title screen with the text "Understanding Bus Bunching"
- **Narration:** "Have you ever waited at a bus stop, only to have two or more buses arrive at the same time? This phenomenon is known as bus bunching, and it's a common issue in public transportation systems around the world."

Scene 2: What is Bus Bunching? (0:30 - 1:00)

- **Visual:** Animated graphic showing a series of buses on a route, with one bus getting delayed and the others catching up.
- **Narration:** "Bus bunching occurs when buses that are scheduled to arrive at regular intervals end up arriving together. This usually happens when one bus gets delayed, causing passengers to pile up at subsequent stops. The next bus, which arrives on time, then gets delayed due to the increased number of passengers, and the cycle continues."

Scene 3: Causes of Bus Bunching (1:00 - 1:30)

- **Visual:** Split screen showing different causes: traffic congestion, high passenger loads, and traffic signals.
- **Narration:** "Several factors contribute to bus bunching, including traffic congestion, high passenger loads, and traffic signals. Even a small delay can cascade into a larger problem, leading to multiple buses arriving at the same time."

Scene 4: Impact on Passengers (1:30 - 2:00)

- **Visual:** Footage of passengers waiting at a bus stop, looking frustrated as two buses arrive simultaneously.
- **Narration:** "For passengers, bus bunching means longer wait times and overcrowded buses. It can be particularly frustrating when you have to wait longer than expected, only to find that the bus is too full to board."

Scene 5: Solutions to Bus Bunching (2:00 - 2:30)

- **Visual:** Animated graphics showing different solutions: bus lanes, better scheduling, and real-time tracking.
- **Narration:** "To address bus bunching, cities can implement dedicated bus lanes, improve scheduling practices, and use real-time tracking systems to adjust bus intervals dynamically. These measures can help ensure that buses arrive on time and

provide a more reliable service."

Scene 6: Conclusion (2:30 - 3:00)

- **Visual:** Closing screen with text "Together, we can make public transportation better!"
- **Narration:** "Bus bunching is a complex issue, but with the right strategies, we can minimize its impact and make public transportation more efficient for everyone. Thank you for watching!"

IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be followed

1. Double-check your data for accuracy before presenting it.
2. Stick to the 3-minute timeframe to keep your audience's attention.
3. Explain technical terms if necessary.
4. Provide context for the data to help the audience understand its significance
5. Get feedback from peers to improve your story telling approach.

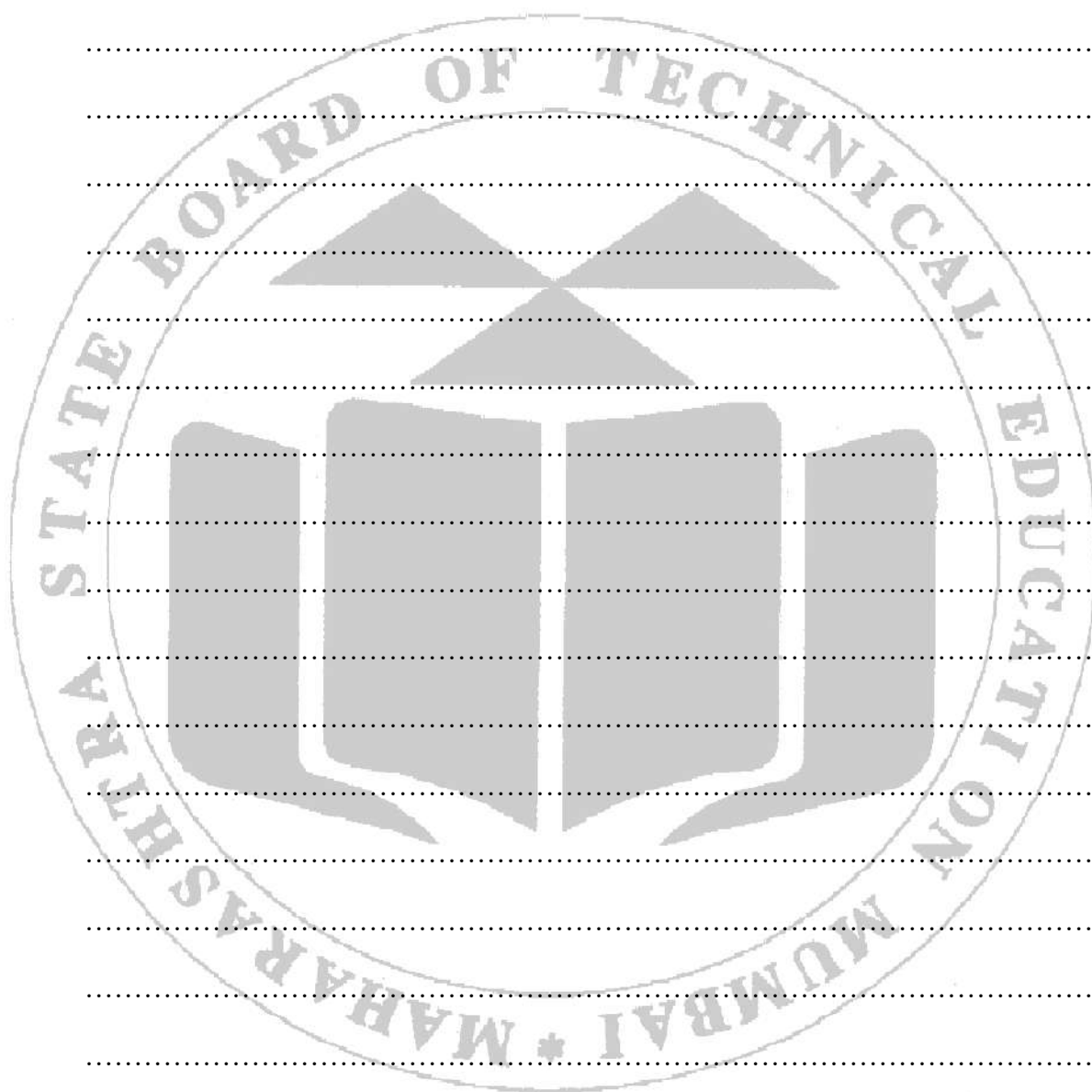
XI Conclusion

XII Practical Related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

- i. Write steps to create the video?
- ii. Give the name of video editing and recording tools.

[Space for Answer]



XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.youtube.com/watch?v=X6Pyc81McJY>
2. https://www.youtube.com/watch?v=6X_YsPketmk
3. <https://www.youtube.com/watch?v=PIMTuf-d49M>
4. <https://www.youtube.com/watch?v=4ivJK77HP2s>
5. <https://www.youtube.com/watch?v=seWYRud-Ch8>

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.6: Create a data story for Vehicle (Bicycle/Bike/Car/Bus etc.) Rental System.

I Practical Significance

From this practical analysis, students will gain hands-on experience in data-driven decision-making within the context of a bike-sharing system. They will learn to identify key usage patterns and trends, understand user demographics and preferences, and evaluate the impact of external factors like weather on service usage. By engaging with real-world data, students will develop skills in data analysis, visualization, and interpretation, enabling them to draw meaningful conclusions that can inform operational and strategic decisions in transportation services and beyond.

II Industry/ Employer Expected Outcome

1. To extract the valuable insights from complex data sets
2. To develop proficiency in creating impactful charts, graphs, and dashboards.
3. To interpret data trends and patterns accurately.

III Course Level Learning Outcome

- CO1- Identify the characters in data Storytelling.
- CO2- Eliminate clutter and grab audience attention.
- CO3- Construct Story telling for the given incident.

IV Laboratory Learning Outcome

- LLO 6.1-Construct a data story on a given incident.
- LLO 6.2-Record the observations of the given data story.

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

The title "Citi Bike Rental Analysis" holds significance as it allows for a detailed examination of key aspects influencing the operation and management of bike-sharing systems. By analyzing popular pick-up locations, understanding trip durations across different age groups and times, and studying user behaviors between subscriber types and days of the week, operators can optimize bike distribution, enhance user experience, and tailor services to meet diverse user needs effectively. Additionally, insights into factors like weather conditions and user demographics provide valuable data for improving service reliability and operational efficiency. Overall, this analysis aids in creating a more efficient and responsive bike-sharing service for urban environments.

VII Exercises

Create a data story on **Uber Ride-Sharing Analysis**.

VIII Procedure with Example:

Data Story: Citi Bike Rental Analysis

Introduction: Citi Bike is a popular bike-sharing system operating in urban areas. This data story explores various aspects of Citi Bike rentals using a dataset collected over a period of time.

Dataset Overview:

- **Variables:** The dataset includes information on pick-up locations, trip durations, user demographics (age), user type (subscriber vs. one-time user), weather conditions, and trip timestamps.
- **Period:** Data spans over several months to capture seasonal and temporal variations.

Key Insights:

1. Most Popular Pick-up Locations:

- **Observation:** Analysis reveals that pick-up locations near major transit hubs, parks, and tourist attractions are the most popular.
- **Impact:** High demand in these areas suggests usage for commuting and ease purposes, influencing station management and bike distribution strategies.

2. Trip Duration across Age Groups and Time:

- **Observation:** Trip durations vary significantly across different age groups and over time.
- **Insight:** Younger adults (ages 20-35) tend to have shorter trip durations, possibly for short commutes or ease rides. Older adults (ages 35-50) may have longer durations, indicating more extended rides or slower cycling speeds.
- **Temporal Variation:** Average trip durations may increase during weekends and evenings, reflecting easy rides or longer exploration trips.

3. Age Group Preferences:

- **Observation:** The 20-35 age group rents bikes the most frequently.
- **Reason:** This age group often values convenience and affordability, using bikes for daily commuting or short-distance travel.

4. User Type and Day of the Week Analysis:

- **Observation:** Long-term subscribers typically use bikes consistently throughout the week, with higher usage during weekdays.
- **One-time Users:** Prefer weekends for easy rides or exploring the city.
- **Impact:** Understanding usage patterns helps in optimizing bike availability and maintenance schedules.

5. Impact of Weather and User Age on Trip Duration:

- **Observation:** Weather conditions (rainy days, temperature) and user age influence average trip duration.
- **Insight:** Adverse weather conditions may shorten trip durations due to decreased comfort and safety concerns. Younger users may tolerate weather variations better than older users.

IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be followed

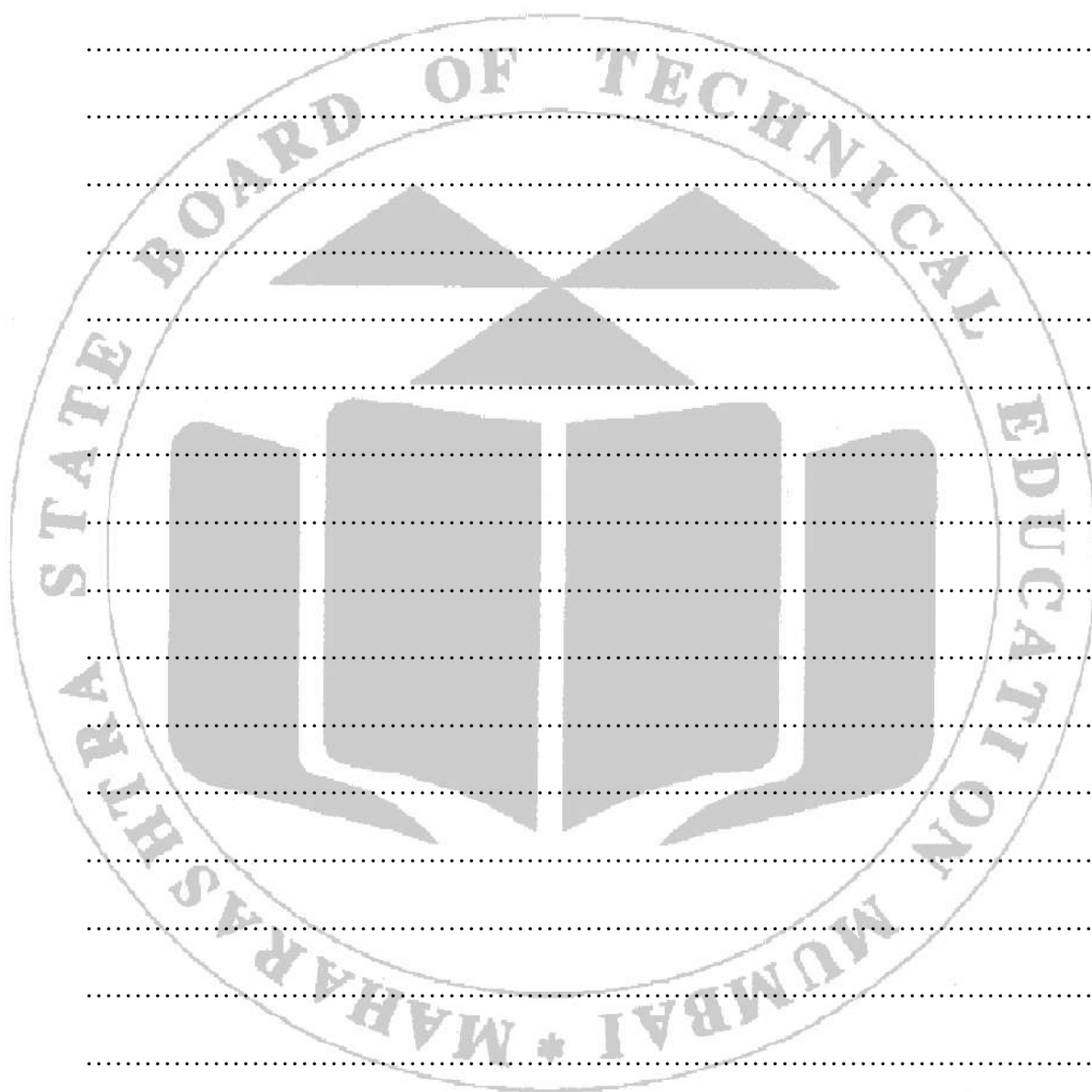
1. Data Integrity: Ensure the accuracy and reliability of the data used for visualization to avoid misleading insights or conclusions.
2. Ethical Representation: Present data and visualizations ethically, avoiding distortion or manipulation that could lead to misinterpretation or biased narratives.

XI Conclusion**XII Practical Related Questions**

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

- What is dataset?
- How the dataset is important in data Storytelling?
- Give the name of the dataset used for analysis.

[Space for Answer]



XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.youtube.com/watch?v=X6Pyc81McJY>
2. <https://www.youtube.com/watch?v=ZFGG2ppSlao>
3. <https://www.youtube.com/watch?v=PIMTuf-d49M>
4. <https://www.youtube.com/watch?v=yQoHcdnG-Yw>

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.7: Create a video (up to 5 minutes) telling a story on given Incidence. Record a video of yourself speaking, or narrate while showing visual props or sketches, or screencast a PowerPoint presentation, etc. Choose how you present the story. Produce a single video file (formatted as a .mov file)

I Practical Significance

From a student's perspective, creating a video to narrate a story on a given incident offers invaluable practical benefits. It allows students to develop essential communication skills by articulating ideas clearly and engagingly. Through the process of planning, scripting, and recording, students enhance their creativity and critical thinking as they decide how best to present information visually and verbally. This practical exercise also fosters technical proficiency with tools like video editing software and presentation platforms, preparing students for future academic and professional endeavors where effective multimedia communication is increasingly important.

II Industry/ Employer Expected Outcome

1. To deliver compelling presentations of data findings.
2. To develop proficiency in creating impactful charts, graphs, and dashboards.
3. To interpret data trends and patterns accurately.

III Course Level Learning Outcome

- CO1-Identify the characters in data Story telling.
- CO2-Eliminate clutter and grab audience attention.
- CO3-Construct Story telling for the given incident.

IV Laboratory Learning Outcome

LLO 7.1-Narrate the data story by compositing a video consisting of various visual props.

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

The creating such a video lies in the integration of Storytelling and multimedia communication. Effective Story telling involves structuring information coherently, selecting key details to emphasize, and engaging the audience through narrative techniques. Understanding the basics of narrative structure, visual communication principles, and audience engagement strategies enriches the student's ability to convey information effectively in diverse contexts. Moreover, exploring different presentation formats—such as narrating over visual props, sketches, or slides— demonstrates how to adapt content for different mediums, enhancing overall communication versatility and impact.

VII Exercises

Create a data story video on **historic event** (e.g. first moon landing)

VIII Procedure

Creating a video with a narrative involves several key steps:

1. **Scriptwriting:** Write a script outlining your story, including an introduction, main plot points, and a conclusion.
2. **Storyboarding:** Plan visuals and sequences using sketches or a storyboard to map out how the story will unfold visually.
3. **Recording:** Record your narration, either with a camera or using screen recording software if you're presenting slides or visuals.
4. **Visuals and Editing:** Gather or create visuals, animations, or slides to accompany your narration. Edit these together with your recorded audio to create a cohesive story.
5. **Finalization:** Review and edit your video for clarity, pacing, and coherence. Add any necessary transitions, music, or effects to enhance the Storytelling.
6. **Exporting:** Export your video file in the desired format (.mov in this case) ensuring it meets any specified requirements for quality and size.

This structured approach will help ensure your video effectively communicates the intended story in an engaging and professional manner.

IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be followed

1. Double-check your data for accuracy before presenting it.
2. Stick to the 5-minute timeframe to keep your audience's attention.
3. Explain technical terms if necessary.
4. Provide context for the data to help the audience understand its significance
5. Get feedback from peers to improve your Story telling approach.

XI Conclusion

.....

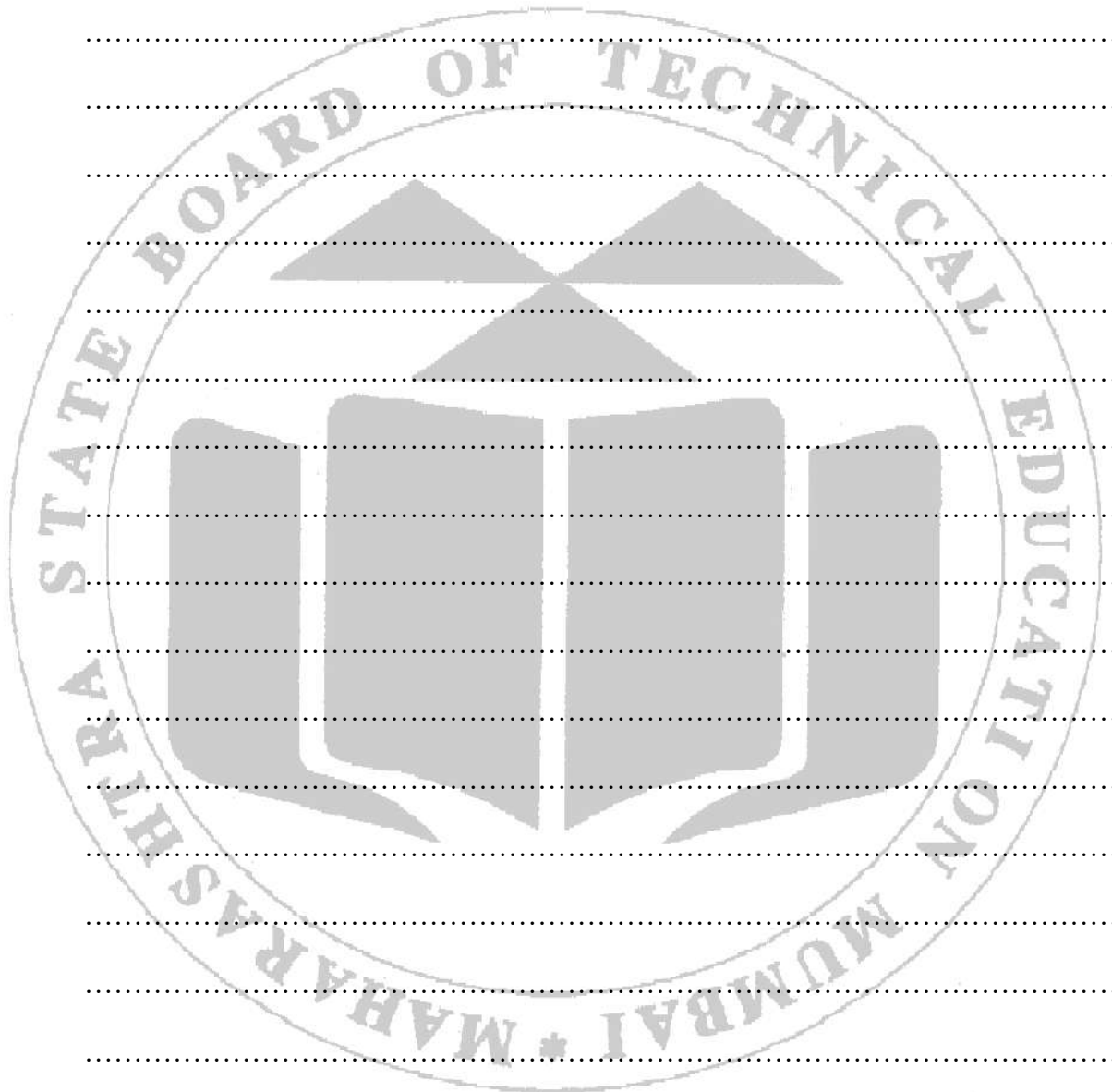
.....

XII Practical Related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

- i. How to export the video in **.mov** file?
- ii. Write the key steps to create the video?
- iii. Give the name of video editing and recording tools.
- iv. Why to add various visual props and sketches in video?

[Space for Answer]



XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.youtube.com/watch?v=4ivJK77HP2s>
2. <https://www.youtube.com/watch?v=seWYRud-Ch8>
3. <https://www.youtube.com/watch?v=vPxyV9kcUig>

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.8: Create a bar chart for data visualizations on Practical No. 6.

I Practical Significance

A bar plot is a common type of visualization used in data analysis to display the distribution of Categorical data. In a bar plot, the categories are typically displayed on the x-axis, while the height of each bar represents the frequency or proportion of observations in each category. The significance of a bar plot depends on the context and the purpose of the analysis. Plots enable us to visualize data in a pictorial or graphical representation.

II Industry/ Employer Expected Outcome

1. To deliver compelling presentations of data findings.
2. To develop proficiency in creating impactful charts, graphs, and dashboards.
3. To use of tools like Tableau, Power BI, or Python for visualization.

III Course Level Learning Outcome

- CO1-Identify the characters in data Story telling.
- CO2-Eliminate clutter and grab audience attention.
- CO3-Construct Story telling for the given incident.

IV Laboratory Learning Outcome

- LLO 8.1-Create a bar chart from the recorded data.

V Relevant Affective domain related Outcome(s)

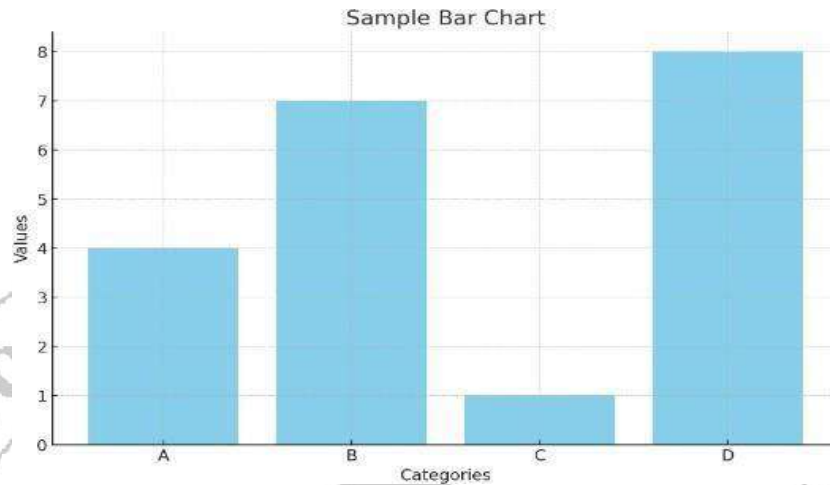
1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

To create a bar chart, start by gathering data and identifying the variables you want to display on the x-axis and y-axis. Next, choose a plotting library such as Matplotlib for Python, and set up your environment by installing the library (e.g., `pip install matplotlib`) and importing it (`import matplotlib.pyplot as plt`). Prepare your data by organizing it into lists, arrays, or dataframes. For instance, you might have `categories = ['A', 'B', 'C', 'D']` and `values = [4, 7, 1, 8]`. Create the bar chart using `plt.bar(categories, values)`. Customize your chart by adding titles (`plt.title('Title of the Bar Chart')`), labels (`plt.xlabel('Categories')`, `plt.ylabel('Values')`), and changing colors (`plt.bar(categories, values, color='skyblue')`). Optionally, add gridlines (`plt.grid(True)`) and legends (`plt.legend(['Legend Label'])`). Finally, display the chart with `plt.show()` or save it as an image file using `plt.savefig('barchart.png')`. Here's a complete example:

```
import matplotlib.pyplot as plt
categories = ['A', 'B', 'C', 'D']
values = [4, 7, 1, 8]
```

```
plt.bar(categories, values, color='skyblue')
plt.title('Sample Bar Chart')
plt.xlabel('Categories')
plt.ylabel('Values')
plt.show()
```

Output:

These steps provide a streamlined process for creating bar charts using Matplotlib in Python.

VII Exercises

Create a bar plot for the following data:

- Average Trip Duration across Age Groups and Over Time:**
 - Age Groups: <20, 20-30, 30-40, 40-50, >50.
 - Average trip duration in minutes (over time):
 - 2022: [12, 15, 18, 17, 14].
 - 2023: [14, 16, 17, 18, 15].
- Age Group Renting the Most Bikes:**
 - Age Groups: <20, 20-30, 30-40, 40-50, >50.
 - Number of rentals: [3000, 8000, 6000, 4000, 2000].

VIII Procedure with Example

To create a meaningful data story and accompanying bar charts based on the observations listed, we'll need to assume some data. Here's an outline of the data we'll use for each observation:

1. Most Popular Pick-up Locations:

- Locations: Central Park, Times Square, Wall Street, Brooklyn Bridge, Empire State Building.
- Number of rentals: [5000, 4500, 3500, 3000, 2500].

Code:

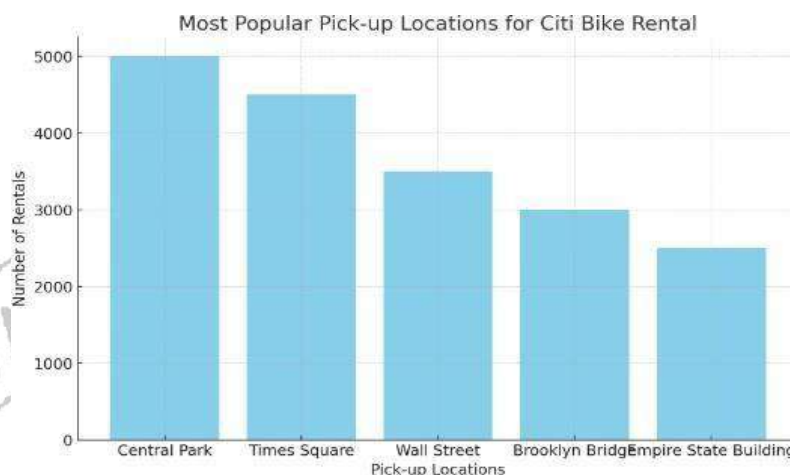
```
import matplotlib.pyplot as plt
# Data
locations = ['Central Park', 'Times Square', 'Wall Street', 'Brooklyn Bridge',
            'EmpireState Building']

rentals = [5000, 4500, 3500, 3000, 2500]

# Bar Chart

plt.figure(figsize=(10, 6))
```

```
plt.bar(locations, rentals, color='skyblue')
plt.xlabel('Pick-up Locations')
plt.ylabel('Number of Rentals')
plt.title('Most Popular Pick-up Locations for Citi Bike Rental')
plt.show()
```

Output:**IX Required Resources**

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be followed

1. Double-check your data for accuracy before presenting it.
2. Explain technical terms if necessary.
3. Provide context for the data to help the audience understand its significance

XI Conclusion

.....

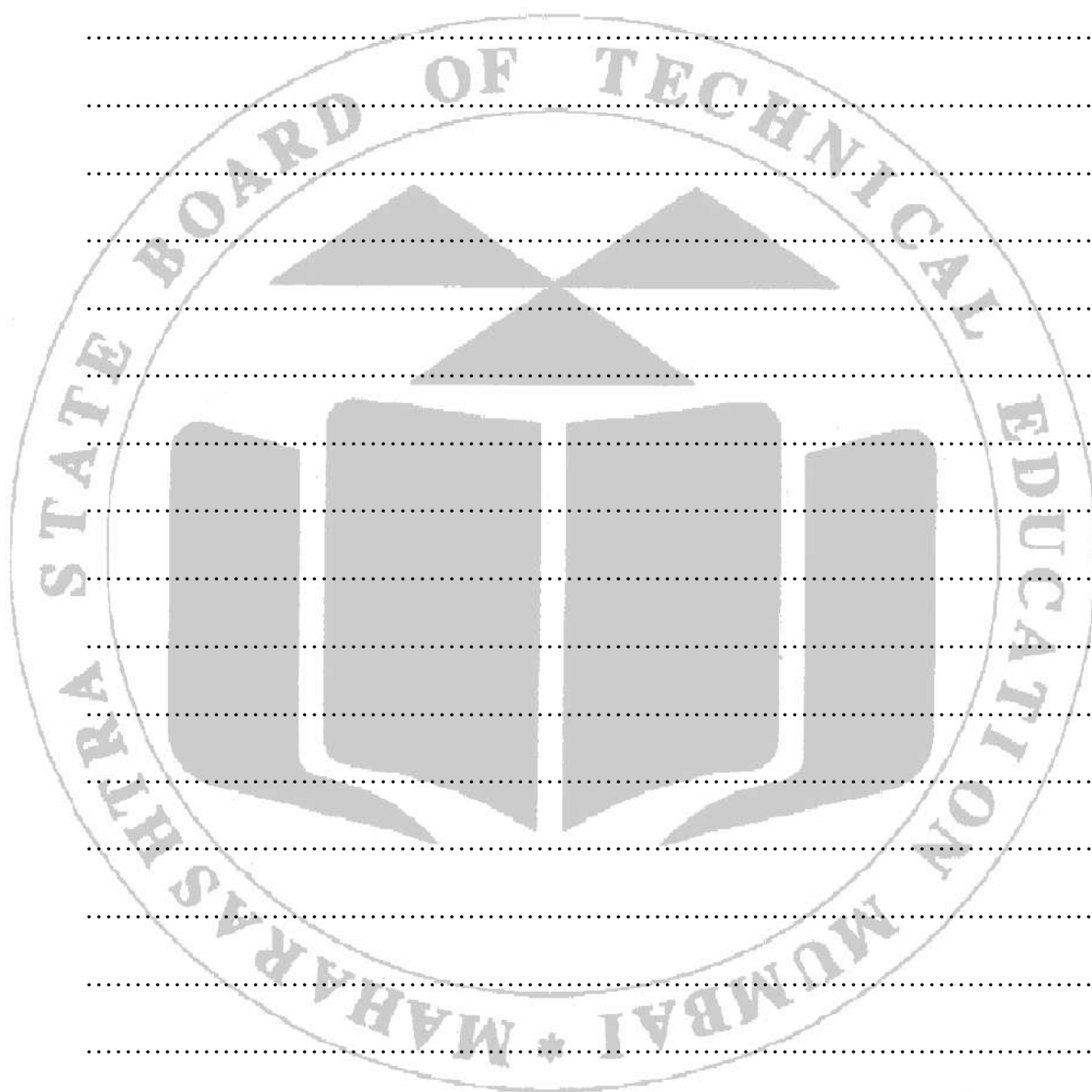
.....

XII Practical Related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

- i. What is data visualization?
- ii. How to plot the bar chart?
- iii. What is **matplotlib.pyplot**?
- iv. How to install matplotlib?

[Space for Answer]



XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.geeksforgeeks.org/bar-plot-in-matplotlib/>
2. https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.bar.html
3. <https://www.dataquest.io/blog/how-to-plot-a-bar-graph-matplotlib/>

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.9: Construct a Case study on data Story telling for any Musical/Social App.

I Practical Significance

Data Story telling for musical or social apps, like Spotify's "Wrapped," enhances user engagement and retention by transforming listening habits into personalized, visually appealing narratives. This approach makes complex data accessible and engaging, encouraging users to interact more with the app and share their experiences on social media. The resulting organic marketing boosts the app's visibility and attracts new users. Additionally, continuous delivery of unique insights fosters a sense of loyalty, keeping users engaged over time.

II Industry/ Employer Expected Outcome

1. To deliver compelling presentations of data findings.
2. To develop proficiency in creating impactful charts, graphs, and dashboards.
3. To use of tools like Tableau, Power BI, or Python for visualization.

III Course Level Learning Outcome

- CO2-Eliminate clutter and grab audience attention.
- CO3-Construct Story telling for the given incident.
- CO4-Transform Data to Visuals.

IV Laboratory Learning Outcome

- LLO 9.1-Explore a real-life example of Storytelling of any musical/social app.

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

Understanding data collection, analysis, visualization, and personalization is essential for effective data Story telling. Collecting user interaction data and analyzing it to identify trends forms the foundation. Visualizing this data through charts and info graphics makes it more understandable. Crafting a compelling narrative that incorporates these insights ensures relatability and engagement. Tailoring stories to individual users using personalization algorithms enhances the relevance and impact of the data stories.

VII Exercises

- Write the steps to create the Case study on data Story telling for "Spotify" Musical App.
- Write a python code to show pie plot data from an excel file in Matplotlib.

VIII Procedure for Constructing a Case Study on Data Story telling for a Musical/Social App:

Step 1: Define Objectives

- **Identify Goals:** Determine the purpose of the case study (e.g., improving user engagement, enhancing user experience, increasing social media shares).
- **Select Metrics:** Decide on the key metrics you want to highlight (e.g., top songs, most-listened artists, genres, listening habits over time).

Step 2: Data Collection

- **Gather Data:** Collect data on user interactions with the app, such as song plays, duration, frequency, likes, shares, and user demographics.
- **Ensure Privacy:** Make sure the data collection process complies with privacy regulations and protects user data.

Step 3: Data Analysis

- **Clean Data:** Preprocess the data to remove any inconsistencies or errors.
- **Analyze Data:** Use statistical and analytical tools to identify trends and patterns in the collected data (e.g., most popular songs, average listening time, favorite genres).
- **Segment Users:** Categorize users into segments based on their listening habits and preferences.

Step 4: Data Visualization

- **Select Tools:** Choose appropriate data visualization tools and software (e.g., Tableau, Power BI, or custom-built solutions).
- **Create Visuals:** Develop engaging and easy-to-understand visuals, such as bar charts, pie charts, line graphs, and info graphics to represent the analyzed data.
- **Design Layout:** Arrange visuals in a logical sequence to form an organized story.

Step 5: Narrative Construction

- **Craft the Story:** Develop a narrative around the data insights, highlighting significant trends, fun facts, and milestones in a way that resonates with users.
- **Personalization:** Tailor the narrative to individual users by incorporating personalized elements based on their specific data.

Step 6: Integration and Testing

- **Integrate in App:** Embed the data stories within the app's user interface, ensuring a seamless and interactive experience.
- **User Testing:** Conduct user testing to gather feedback and make necessary adjustments to improve the Story telling experience.

Step 7: Launch and Promotion

- **Launch Feature:** Release the data Story telling feature to the user base.
- **Promote on Social Media:** Encourage users to share their personalized data stories on social media platforms to boost visibility and engagement.

Step 8: Monitor and Iterate

- **Collect Feedback:** Gather user feedback and analyze engagement metrics post-launch.
- **Iterate and Improve:** Continuously refine the data Story telling feature based on feedback and new data insights to enhance user experience and maintain engagement.

Sample Python code:

```
import pandas as pd
import matplotlib.pyplot as plt
```


Sample data as a dictionary

```
data_dict = {
    'user_id': ['user_1', 'user_1', 'user_1', 'user_1', 'user_1', 'user_2', 'user_2'],
    'song_id': ['song_1', 'song_2', 'song_3', 'song_4', 'song_5', 'song_6', 'song_7'],
    'artist': ['artist_A', 'artist_B', 'artist_A', 'artist_C', 'artist_D', 'artist_E', 'artist_F'],
    'genre': ['Pop', 'Rock', 'Pop', 'Hip-Hop', 'Jazz', 'Classical', 'Pop'],
    'timestamp': ['2023-01-01 12:00:00', '2023-01-02 13:00:00', '2023-01-03
14:00:00', '2023-01-04 15:00:00', '2023-01-05 16:00:00', '2023-01-01
17:00:00', '2023-01-02 18:00:00']
}
```

Convert the dictionary to a DataFrame

```
data = pd.DataFrame(data_dict)
```

Convert timestamp to datetime

```
data['timestamp'] = pd.to_datetime(data['timestamp'])
```

Extract year and month for analysis

```
data['year'] = data['timestamp'].dt.year
```

```
data['month'] = data['timestamp'].dt.month
```

Display the first few rows of the dataset

```
print("Dataset:")
```

```
print(data)
```

Calculate top songs for each user

```
top_songs = data.groupby(['user_id', 'song_id']).size().reset_index(name='count')
```

```
top_songs = top_songs.sort_values(['user_id', 'count'], ascending=[True, False])
```

```
top_songs = top_songs.groupby('user_id').head(5)
```

```
print("\nTop Songs for Each User:")
```

```
print(top_songs)
```

Calculate top artists for each user

```
top_artists = data.groupby(['user_id', 'artist']).size().reset_index(name='count')
```

```
top_artists = top_artists.sort_values(['user_id', 'count'], ascending=[True, False])
```

```
top_artists = top_artists.groupby('user_id').head(5)
```

```
print("\nTop Artists for Each User:")
```

```
print(top_artists)
```

Calculate top genres for each user

```
top_genres = data.groupby(['user_id', 'genre']).size().reset_index(name='count')
```

```
top_genres = top_genres.sort_values(['user_id', 'count'], ascending=[True, False])
```

```
top_genres = top_genres.groupby('user_id').head(5)
```

```
print("\nTop Genres for Each User :")
```

```
print(top_genres)
```

Visualization for a specific user

```
user_id = 'user_1'
```

Filter data for the selected user

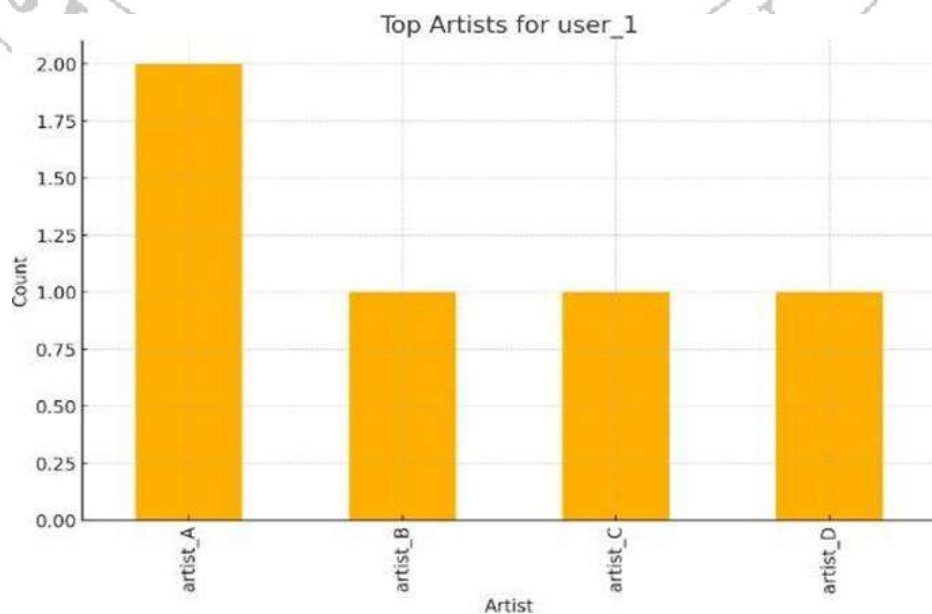
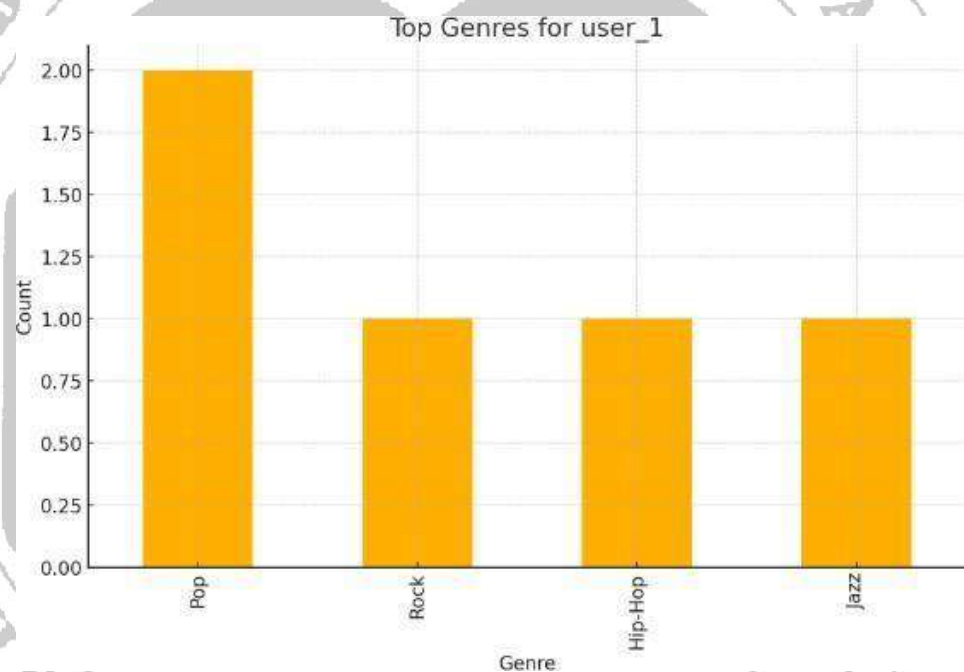
```
user_data = data [data ['user_id'] == user_id]
```

Calculate the top genres for the user

```
user_top_genres = user_data['genre'].value_counts().head(5)
user_top_genres.plot(kind='bar', title=f'Top Genres for {user_id}',
xlabel='Genre',ylabel='Count')
plt.show()
```

Calculate the top artists for the user

```
user_top_artists = user_data['artist'].value_counts().head(5)
user_top_artists.plot(kind='bar', title=f'Top Artists for {user_id}',
xlabel='Artist',ylabel='Count')
plt.show()
```

Output:

IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be followed

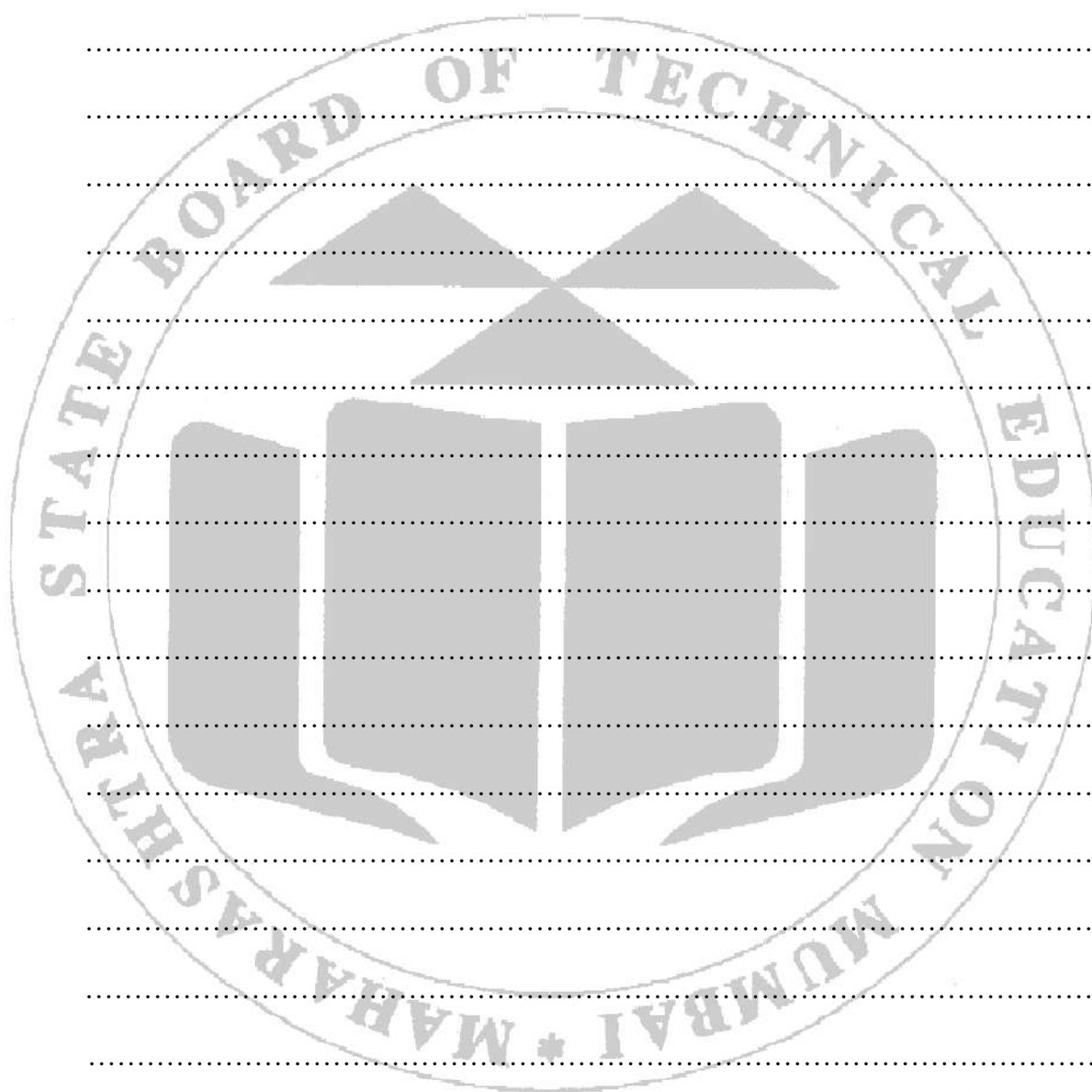
1. Double-check your data for accuracy before presenting it.
2. Explain technical terms if necessary.
3. Provide context for the data to help the audience understand its significance.

XI Conclusion**XII Practical Related Questions**

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

- Define case study?
- How to plot the pie chart?
- What are the other data visualization tools available like **matplotlib**?
- How to install pandas?

[Space for Answer]



XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.geeksforgeeks.org/bar-plot-in-matplotlib/>
2. https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.bar.html
3. <https://www.geeksforgeeks.org/plot-data-from-excel-file-in-matplotlib-python/>
4. <https://www.tutorialspoint.com/how-to-plot-pie-chart-with-a-single-pie-highlighted-with-python-matplotlib>

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.10: Implementation of a python program that performs data cleaning on any dataset.

I Practical Significance

Data cleaning is practically significant because the quality of data directly impacts the accuracy and reliability of the analysis, machine learning models, and decision-making processes that rely on it. Clean data helps in reducing biases and errors in statistical conclusions and predictions, leading to more robust and actionable insights. For businesses, this can mean better customer insights, improved operational efficiency, and more informed strategic decisions. In research, it ensures that findings are valid and replicable. Overall, clean data is foundational to any data-driven endeavor, ensuring that subsequent analyses are based on reliable information.

II Industry/ Employer Expected Outcome

1. To deliver compelling presentations of data findings.
2. To develop proficiency in creating impactful charts, graphs, and dashboards.
3. To use of tools like Tableau, Power BI, or Python for visualization.

III Course Level Learning Outcome

- CO2-Eliminate clutter and grab audience attention.
- CO3-Construct Story telling for the given incident.
- CO4-Transform Data to Visuals.

IV Laboratory Learning Outcome

LLO 10.1-Write a Python program for data cleaning

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

Data cleaning, also known as data cleansing or data preprocessing, is a critical step in the data preparation process. It involves identifying and correcting (or removing) errors and inconsistencies in data to improve data quality. Common tasks include handling missing values, removing duplicates, and addressing outliers. The objective is to ensure the data is accurate, complete, and suitable for analysis. This process is essential because data often comes from various sources and can contain errors due to manual entry, merging from different datasets, or other issues.

VII Exercises

1. Implementation of a python program that performs Data Cleaning on any dataset.
2. Write a python program to discover and remove the duplicate from the file.
3. Write a python program to calculate the MEAN and replace any empty values with it using pandas.

VIII Procedure/Steps for Data Cleaning in Python:

1. **Load the Dataset:** Read the dataset into a Data Frame.
2. **Identify Missing Values:** Check for missing data in the dataset.
3. **Handle Missing Values:** Decide and implement a strategy to fill or remove missing values.
4. **Identify and Handle Duplicates:** Detect and remove duplicate entries in the dataset.
5. **Correct Data Types:** Ensure data types are appropriate for each column.
6. **Address Outliers:** Identify and handle outliers in numerical columns if necessary.
7. **Normalize or Standardize Data:** Scale numerical data if needed for analysis or modeling.
8. **Feature Engineering:** Create new features from existing data if beneficial.
9. **Save Cleaned Dataset:** Export the cleaned dataset for further analysis or modeling.

These steps provide a structured approach to cleaning and preparing a dataset using Python, ensuring it is ready for analysis or machine learning tasks.

Sample Code:

```
import pandas as pd
from sklearn.preprocessing import StandardScaler
# Step 1: Load the dataset
df = pd.read_csv('your_dataset.csv')
# Step 2: Check for missing values
print("Missing values before cleaning:")
print(df.isnull().sum())
# Step 3: Handle missing values
df.fillna(df.mean(), inplace=True)
# Step 4: Remove duplicates
df.drop_duplicates(inplace=True)
# Step 5: Convert data types if needed
df['numeric_column'] = pd.to_numeric(df['numeric_column'])
# Step 6: Handle outliers (using IQR)
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1
df = df[~((df < (Q1 - 1.5 * IQR)) | (df > (Q3 + 1.5 * IQR))).any(axis=1)]
# Step 7: Standardize numeric data
scaler = StandardScaler()
df[['numeric_column1', 'numeric_column2']] =
scaler.fit_transform(df[['numeric_column1', 'numeric_column2']])
# Step 8: Save cleaned dataset
df.to_csv('cleaned_dataset.csv', index=False)
print("Data cleaning completed. Cleaned dataset saved.")
```

IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be followed

1. Double-check your data for accuracy before presenting it.
2. Explain technical terms if necessary.
3. Provide context for the data to help the audience understand its significance.
4. Make sure to replace "dataset.csv" with your actual dataset's filename.

XI Conclusion**XII Practical Related Questions**

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. What is data cleaning? How can it be done in python?
2. What's the difference between categorical data and continuous data?

[Space for Answer]

XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.datacamp.com/tutorial/guide-to-data-cleaning-in-python>
2. <https://www.geeksforgeeks.org/plot-data-from-excel-file-in-matplotlib-python/>
3. <https://monkeylearn.com/blog/data-cleaning-python/>
4. https://www.youtube.com/watch?v=OS2m0f2gVJ0&ab_channel=SoumilShah
5. https://www.youtube.com/watch?v=iaZQF8SLHJs&ab_channel=RyanNolanData

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.11: Create Bar chart for data visualization using Single distribution. (e. g. Histogram of the gender and ages of the train/flight passengers.)

I Practical Significance

Visualizing the age distribution of train or flight passengers is crucial for several practical reasons. Firstly, it helps transportation companies understand their customer demographics better, enabling them to tailor services and amenities accordingly. For instance, if a significant portion of passengers are elderly, accessibility and comfort might be prioritized differently compared to a younger demographic where technology and connectivity might be more valued. Secondly, understanding age distribution aids in planning marketing strategies. For example, targeting promotions or discounts based on age groups can be more effective when backed by data on passenger age demographics.

II Industry/ Employer Expected Outcome

1. To deliver compelling presentations of data findings.
2. To interpret data trends and patterns accurately.
3. To develop proficiency in creating impactful charts, graphs, and dashboards.
4. To use of tools like Tableau, Power BI, or Python for visualization.

III Course Level Learning Outcome

CO4-Transform Data to Visuals.

CO5-Create data visualization using many distributions

IV Laboratory Learning Outcome

LLO 11.1-Plot histogram using single distribution.

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

The histogram, which is essentially a bar chart representing a single distribution, is rooted in statistical principles of data visualization. It utilizes binning techniques to categorize continuous data (like ages) into discrete intervals or bins, making it easier to interpret patterns and trends within the data. The choice of bin width or interval is critical as it influences how the distribution appears; too wide a bin might smooth out important details, while too narrow a bin might obscure overall trends. Histograms are fundamental in exploratory data analysis (EDA) and are widely used across various fields including transportation logistics, sociology, and market research for their simplicity and effectiveness in conveying distributional information.

To create a bar chart for visualizing a single distribution, such as the ages of train or flight passengers, you would typically organize the data into age groups (**bins**) and count

the number of passengers falling into each group. Each bar in the chart represents a range of ages (e.g., 0-10, 11- 20, 21-30, etc.) on the x-axis, and the height of each bar corresponds to the frequency or count of passengers within that age range on the y-axis. This visualization allows you to quickly grasp the distribution of ages among passengers, identifying trends like age demographics and potential outliers.

VII Exercises

Implementation of a python program create Bar chart for data visualization using Single distribution.

VIII Procedure/Steps to plot Histograms in Python:Explanation of Each Step:

1. **Installation:** Ensure that the necessary library (matplotlib) is installed.
2. **Import:** Import `matplotlib.pyplot` as `plt` to access plotting functions.
3. **Data Preparation:** Define your dataset. This can be from a list, a file, or other data sources.
4. **Plotting:** Use `plt.hist()` to create the histogram. The `bins` parameter controls the number of bins.
5. **Customization:** Add a title and axis labels using `plt.title()`, `plt.xlabel()`, and `plt.ylabel()`.
6. **Display:** Use `plt.show()` to render and display the histogram.

Sample Code:

```
import matplotlib.pyplot as plt

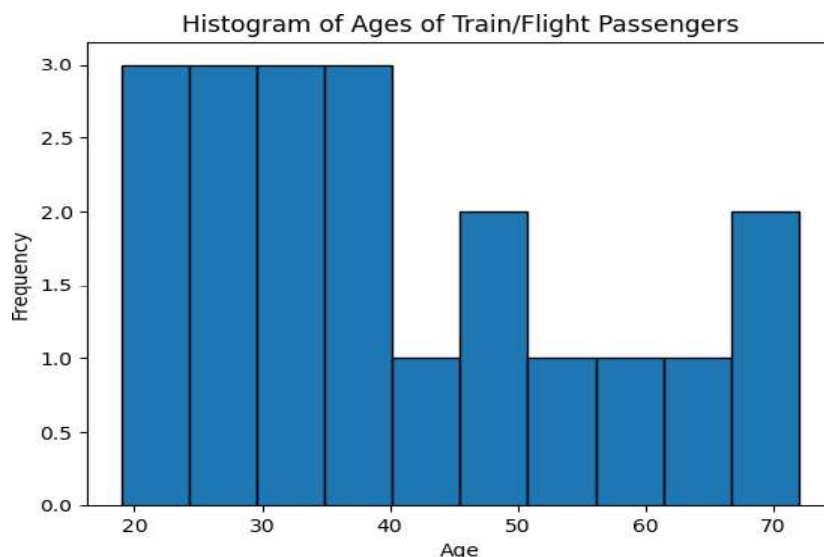
# Sample data
ages = [23, 45, 56, 22, 34, 36, 29, 40, 65, 70, 34, 28, 33, 25, 19, 50, 60, 72, 48, 39]

# Create histogram
plt.hist(ages, bins=10, edgecolor='black')

# Add titles and labels
plt.title('Histogram of Ages of Train/Flight Passengers')
plt.xlabel('Age')
plt.ylabel('Frequency')

# Show plot
plt.show()
```

Output:



IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be followed

1. Double-check your data for accuracy before presenting it.
2. Explain technical terms if necessary.
3. Provide context for the data to help the audience understand its significance.

XI Conclusion

.....

.....

XII Practical Related Questions

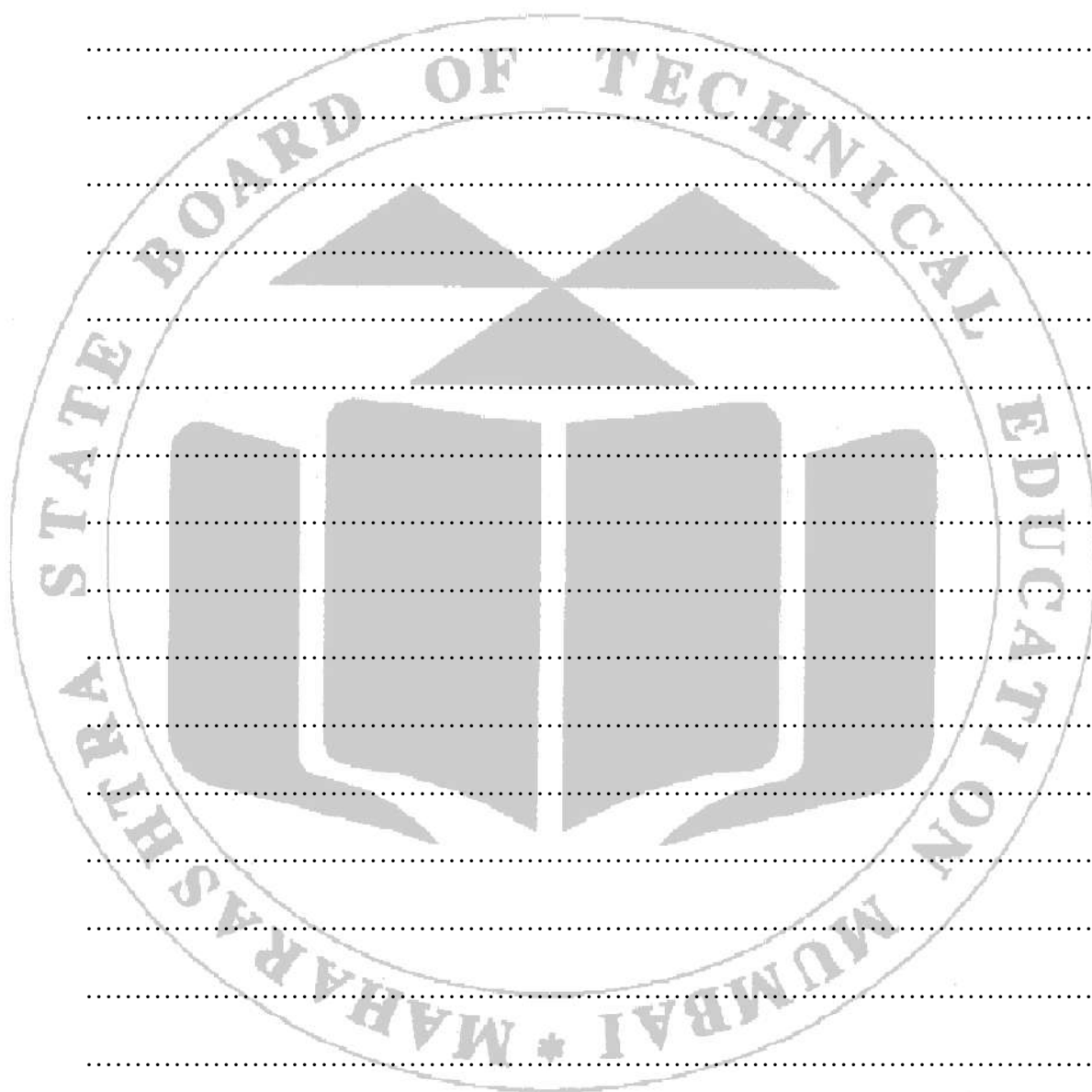
Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. What is single distribution and multiple distribution in data visualization?
2. What is the use of Histograms?

[Space for Answer]

.....

.....



XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics/histograms/v/histograms-intro>
2. <https://www.javatpoint.com/histogram-graph>

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.12: Develop a worksheet, add filters and create chart using dataset by using any Visualization tool.

I Practical Significance

Developing a worksheet, adding filters, and creating charts using a dataset with a visualization tool significantly enhances data analysis and decision-making processes. These tasks enable users to transform raw data into insightful, visually appealing formats, making complex data more understandable. Filters allow users to drill down into specific subsets of data, enabling targeted analysis and better insights. Charts and graphs help in identifying patterns, trends, and outliers, facilitating quicker and more informed decisions. Such skills are crucial in various fields, including business, healthcare, education, and research, where data-driven decisions are essential for efficiency and growth.

II Industry/ Employer Expected Outcome

1. To deliver compelling presentations of data findings.
2. To interpret data trends and patterns accurately.
3. To develop proficiency in creating impactful charts, graphs, and dashboards.
4. To use of tools like Tableau, Power BI, or Python for visualization.

III Course Level Learning Outcome

CO4-Transform Data to Visuals.

CO5-Create data visualization using many distributions

IV Laboratory Learning Outcome

LLO 12.1-Create worksheet using visualization tool.

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

To develop a worksheet, add filters, and create charts using a visualization tool, one must understand basic data concepts such as datasets, variables, and data types (numerical and categorical). Familiarity with spreadsheet functionalities like data entry, sorting, and basic formulas is beneficial. Knowledge of different types of charts (e.g., bar, line, pie) and their appropriate use cases is also essential. Understanding how filters work to segment data without altering the underlying dataset is necessary. Lastly, a basic grasp of how visualization tools like Microsoft Excel, Tableau, or Google Sheets function will enable users to efficiently manipulate and present data.

VII Exercises

1. Implementation of a python program create Bar chart for data visualization using Single distribution.

VIII Procedure/Steps with Example:

- **Be ready with your dataset:** Here we have used following 'students_marks.csv'

	A	B	C
1	student_id	subject	marks
2	1	Math	45
3	2	English	78
4	3	Math	88
5	4	Science	52
6	5	English	69
7	6	Math	95
8	7	Science	45
9	8	English	55
10	9	Math	87
11	10	Science	65

1. **Install the libraries using following commands:**

```
python -m pip install pandas
python -m pip install matplotlib
python -m pip install seaborn
```
2. **Import the above libraries**
3. **Load the Dataset:** Use a visualization tool to import the dataset.
4. **Develop the Worksheet:** Organize and clean the data as needed.
5. **Add Filters:** Apply filters to focus on specific subsets of data.
6. **Create a Chart:** Select the filtered data and generate a chart.
7. **Customize and Save:** Customize the chart with titles and labels, then save your work.

Sample Code:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Load the Dataset
# Replace 'students_marks.csv' with the path to your dataset file.
data = pd.read_csv('students_marks.csv')
# Display the first few rows of the dataset
print(data.head())

# Step 2: Develop the Worksheet
# Clean the data (if necessary)
# For example, drop rows with missing values
data_cleaned = data.dropna()

# Step 3: Add Filters
# apply filters to the dataset
# For example, filter students who scored above 50 in a specific subject
threshold_value = 50
filtered_data = data_cleaned[data_cleaned['marks'] > threshold_value]

# Step 4: Create a Chart
# Create a bar chart showing average marks per subject
```



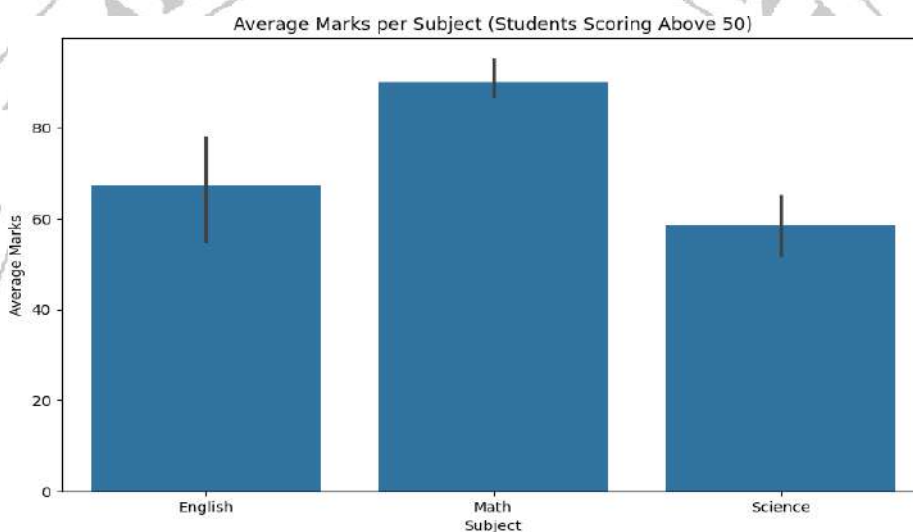
```
plt.figure(figsize=(10, 6))
sns.barplot(x='subject', y='marks', data=filtered_data, estimator='mean')

# Customize the chart
plt.title('Average Marks per Subject (Students Scoring Above 50)')
plt.xlabel('Subject')
plt.ylabel('Average Marks')

# Display the chart
plt.show()

# Step 5: Save the Chart
# Save the chart as an image file
plt.savefig('average_marks_per_subject.png')
```

Output:



IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be followed

1. Double-check your data for accuracy before presenting it.
2. Explain technical terms if necessary.
3. Provide context for the data to help the audience understand its significance.

XI Conclusion

.....

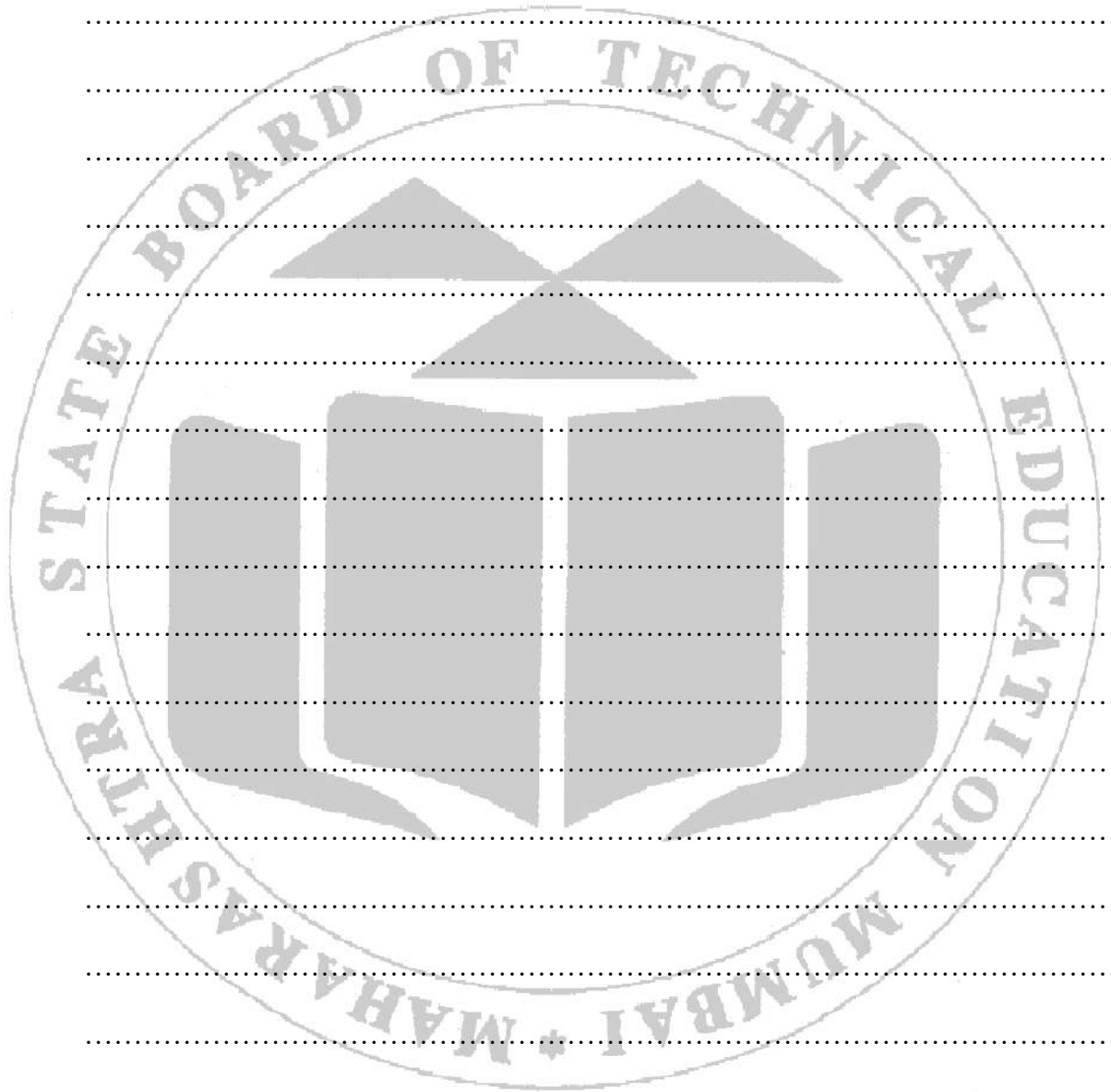
.....

XII Practical Related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. How to install libraries like matplotlib, pandas?
2. How to Install **Openpyxl**?

[Space for Answer]



XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.geeksforgeeks.org/reading-csv-files-in-python/>
2. https://www.w3schools.com/python/pandas/pandas_csv.asp

XIV Assessment Scheme

XV Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.13: Create Bar chart for data visualization using many distribution. (e. g. Histogram of the gender and ages of the train/flight passengers.)

I Practical Significance

Understanding the age distribution of passengers is crucial for various operational aspects in transportation. For airlines, it helps in tailoring services such as in-flight entertainment, meal options, and even seat configurations based on the predominant age groups onboard. For railways, it informs decisions on ticket pricing, seating arrangements, and amenities to cater effectively to different age demographics. Moreover, it aids in planning marketing strategies and customer service initiatives tailored to the specific needs and preferences of different age groups.

II Industry/ Employer Expected Outcome

1. To deliver compelling presentations of data findings.
2. To interpret data trends and patterns accurately.
3. To develop proficiency in creating impactful charts, graphs, and dashboards.
4. To use of tools like Tableau, Power BI, or Python for visualization.

III Course Level Learning Outcome

CO4-Transform Data to Visuals.

CO5-Create data visualization using many distributions

IV Laboratory Learning Outcome

LLO 13.1-Create a histogram using many distribution.

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

XVI Relevant Theoretical Background

Histograms are a fundamental tool in statistical analysis and data visualization. They depict the frequency distribution of a continuous variable (such as age) by grouping data into bins and plotting the number of observations falling into each bin as bars. The width of each bar represents the range of values in each bin, while the height represents the frequency or count of observations. Histograms provide insights into the shape, center, and spread of data distribution, helping analysts and decision-makers to identify patterns, outliers, and general trends within datasets. They are widely used across various disciplines, including economics, sociology, and market research, for exploring and interpreting numerical data effectively.

VI Exercises

1. Implementation python program create Bar chart for data visualization using many distribution.

VII Procedure/Steps to plot Histograms in Python:Explanation of Each Step:

1. **Installation:** Ensure that the necessary library (**matplotlib**) is installed.
2. **Import:** Import **matplotlib.pyplot** as **plt** to access plotting functions.
3. **Data Preparation:** Define your dataset. This can be from a list, a file, or other data sources.
4. **Plotting:** Use **plt.hist()** to create the histogram. The **bins** parameter controls the number of bins.
5. **Customization:** Add a title and axis labels using **plt.title()**, **plt.xlabel()**, and **plt.ylabel()**.
6. **Display:** Use **plt.show()** to render and display the histogram.

Here's an example Python code snippet using Matplotlib and NumPy to visualize and analyze the hypothetical income distributions for doctors, engineers, and artists:

Sample Code:

```
import matplotlib.pyplot as plt
import numpy as np

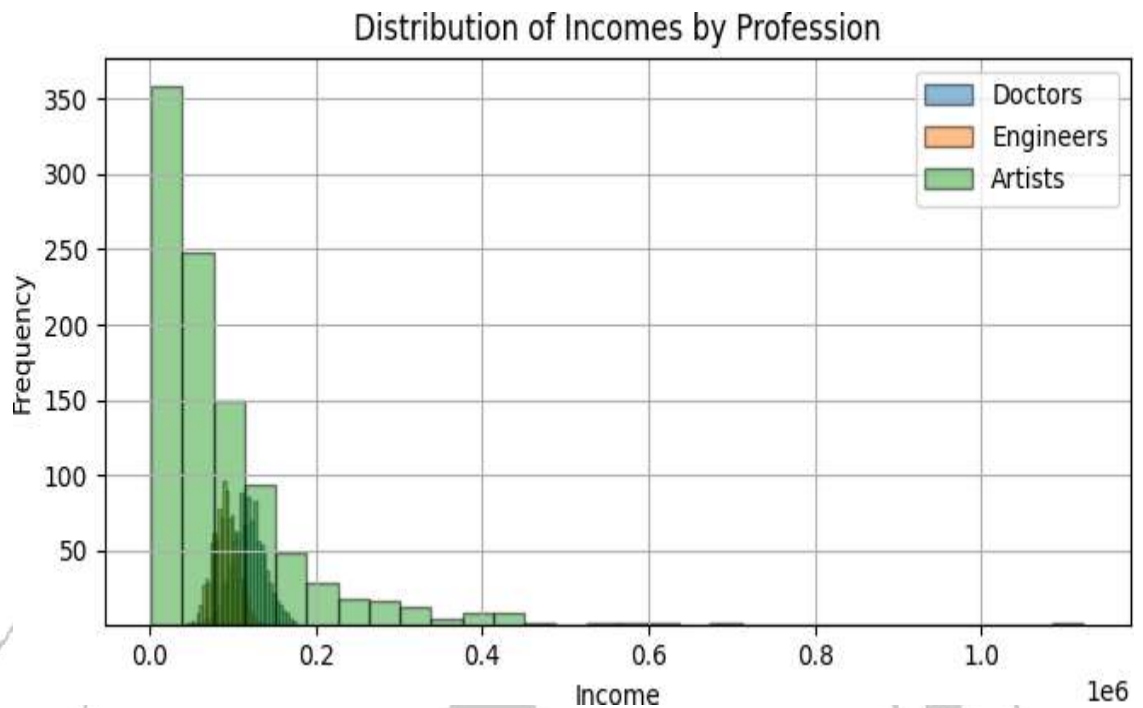
# Generate sample data for incomes
np.random.seed(0)

# Sample data for doctors' incomes (normal distribution with mean 120,000
and standard deviation 20,000)
doctors_incomes = np.random.normal(120000, 20000, 1000)

# Sample data for engineers' incomes (normal distribution with mean 90000
and standard deviation 15000)
engineers_incomes = np.random.normal(90000, 15000, 1000)

# Sample data for artists' incomes (lognormal distribution with mean 60000
and standard deviation 10000)
artists_incomes = np.random.lognormal(np.log(60000), 1, 1000)

# Plotting histograms
plt.figure(figsize=(10, 6))
plt.hist(doctors_incomes, bins=30, alpha=0.5, label='Doctors', edgecolor='black')
plt.hist(engineers_incomes, bins=30, alpha=0.5, label='Engineers',
edgecolor='black')
plt.hist(artists_incomes, bins=30, alpha=0.5, label='Artists', edgecolor='black')
plt.xlabel('Income')
plt.ylabel('Frequency')
plt.title('Distribution of Incomes by Profession')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```

Output:

To plot histograms in Python using Matplotlib, you can follow these steps to visualize the age distribution of train or flight passengers. Here's a step-by-step Python code example:

Sample Code:**# Step 1: Import libraries**

```
import matplotlib.pyplot as plt
import numpy as np
```

Step 2: Generate example data (ages of passengers)**# For reproducibility**

```
np.random.seed(0)
```

Generate 200 random ages between 1 and 100

```
ages = np.random.randint(1, 100, size=200)
```

Step 3: Plot histogram**# Adjust figure size (optional)**

```
plt.figure(figsize=(8, 6))
```

Plot histogram with 20 bins

```
plt.hist(ages, bins=20, color='skyblue', edgecolor='black')
```

```
plt.xlabel('Age')
```

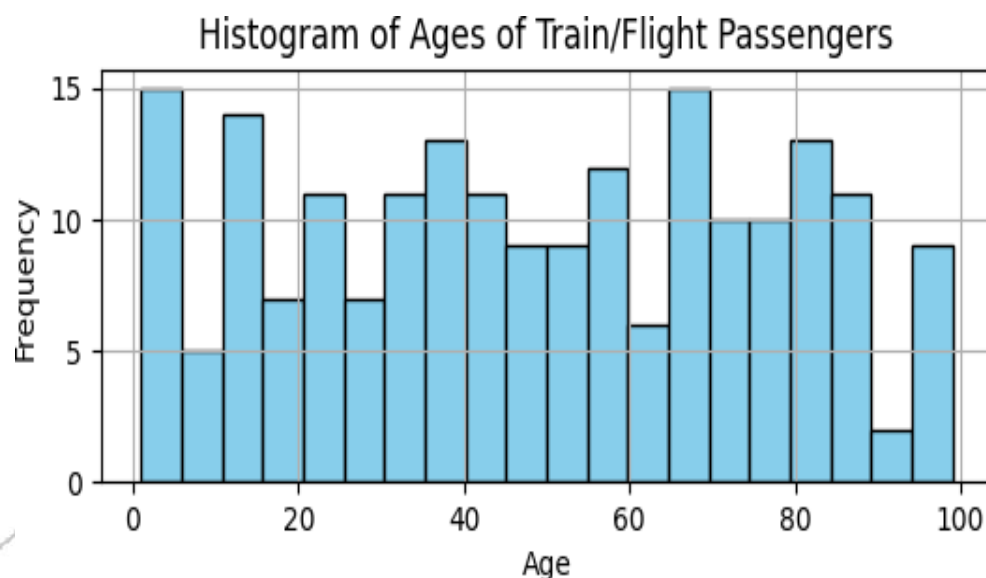
```
plt.ylabel('Frequency')
```

```
plt.title('Histogram of Ages of Train/Flight Passengers')
```

Add grid lines (optional)

```
plt.grid(True)
```

```
plt.show()
```

Output:**VIII Required Resources**

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

IX Precautions to be followed

1. Double-check your data for accuracy before presenting it.
2. Explain technical terms if necessary.
3. Provide context for the data to help the audience understand its significance.

X Conclusion

.....

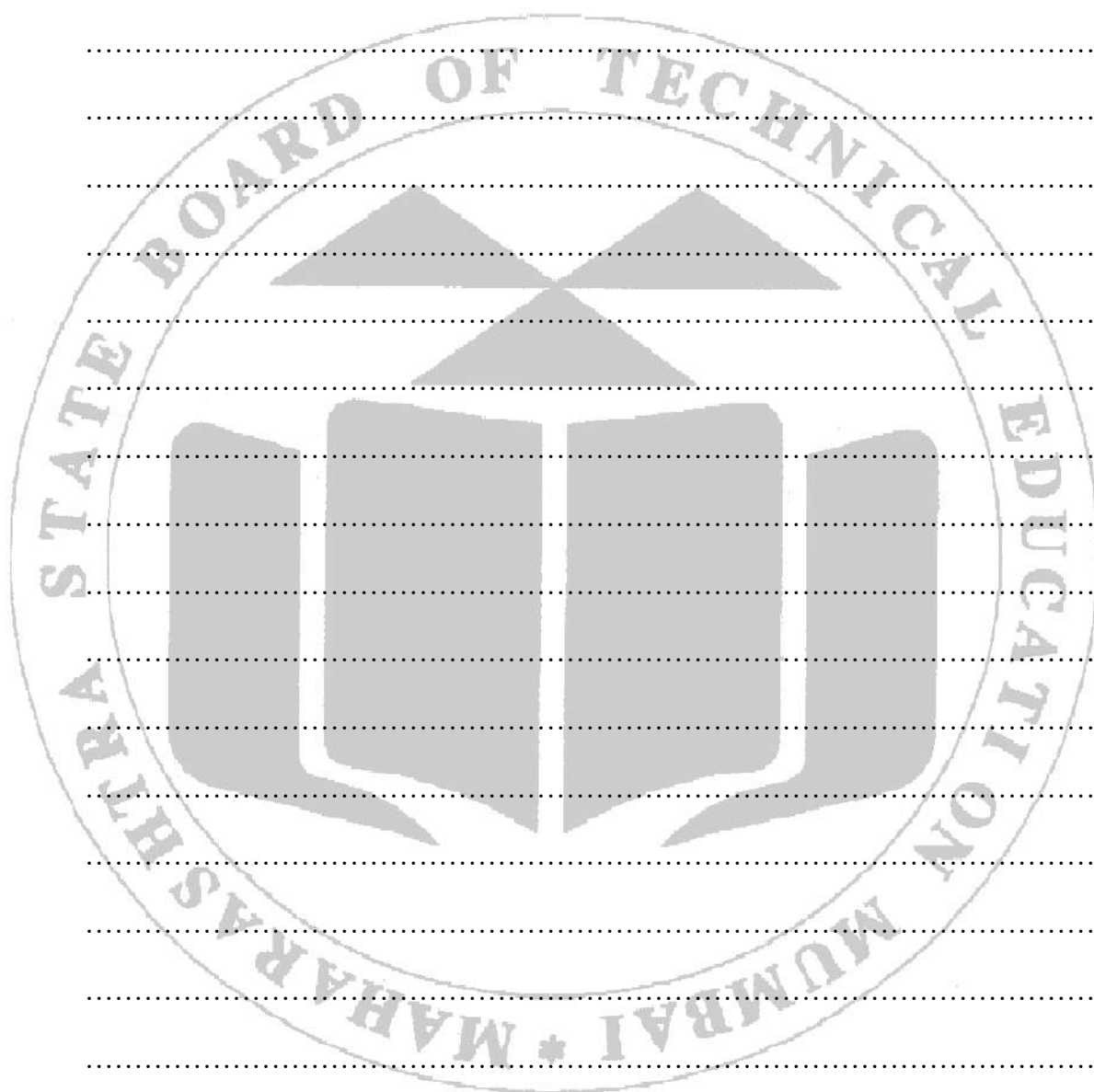
.....

XI Practical Related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. What is difference between single distribution and multiple distribution in data visualization?
2. Write command to install **numpy**?

[Space for Answer]



XII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics/histograms/v/histograms-intro>
2. <https://www.javatpoint.com/histogram-graph>
3. <https://towardsdatascience.com/distribution-visualization-101-with-python-a12d481afec7>

XIII Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.14: Implementation of a python program that loads a dataset and plot grouped bars.

I Practical Significance

Implementing a Python program to load a dataset and plot grouped bars is highly practical across various domains such as data analysis, research, and visualization. This task allows analysts and researchers to efficiently compare multiple categories or groups within a dataset visually. By visualizing data in grouped bars, one can quickly discern patterns, trends, and differences between groups, aiding in decision-making processes. For instance, in business analytics, this could mean comparing sales performance across different regions over time. In academic research, it could involve comparing experimental results across various conditions. The practical significance lies in its ability to simplify complex data into easily interpretable visual insights, facilitating clearer communication and informed decision-making.

II Industry/ Employer Expected Outcome

1. To develop proficiency in creating impactful charts, graphs, and dashboards.
2. To use of tools like Tableau, Power BI, or Python for visualization.

III Course Level Learning Outcome

- CO3-Construct Story telling for the given incident.
- CO4-Transform Data to Visuals.
- CO5-Create data visualization using many distributions.

IV Laboratory Learning Outcome

- LLO14.1-Perform dataset loading to plot grouped bars.
- LLO 14.2-Write a Python program to create grouped bars from the loaded data.

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

To implement a Python program for loading a dataset and plotting grouped bars, a basic understanding of programming fundamentals and key libraries is essential. Knowledge of Python programming language syntax, especially handling data structures like lists and dictionaries, is fundamental. Additionally, familiarity with libraries such as Pandas for data manipulation and Matplotlib or Seaborn for plotting is necessary. Understanding how to read data from CSV files or other formats using Pandas, and then using Matplotlib or Seaborn to create grouped bar plots involves knowing functions like **read_csv()** from Pandas and **bar()** or **barplot()** from Matplotlib or Seaborn respectively. This minimal theoretical background ensures proficiency in data handling and visualization, enabling

effective implementation of the program to load datasets and create grouped bar plots in Python.

VII Exercises

Implement python program create grouped Bar chart for data visualization for any dataset

VIII Procedure/Steps to plot grouped bars in Python:

Here's a sample Python program that demonstrates how to load a dataset from a .CSV file and plot grouped bars using Pandas and Matplotlib libraries:

Code:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

# Sample dataset (you can replace this with your own CSV file)
data = {
    'Category': ['A', 'B', 'C', 'D'], 'Group1': [10, 15, 7, 10], 'Group2': [12, 14, 8, 11]
}
df = pd.DataFrame(data)

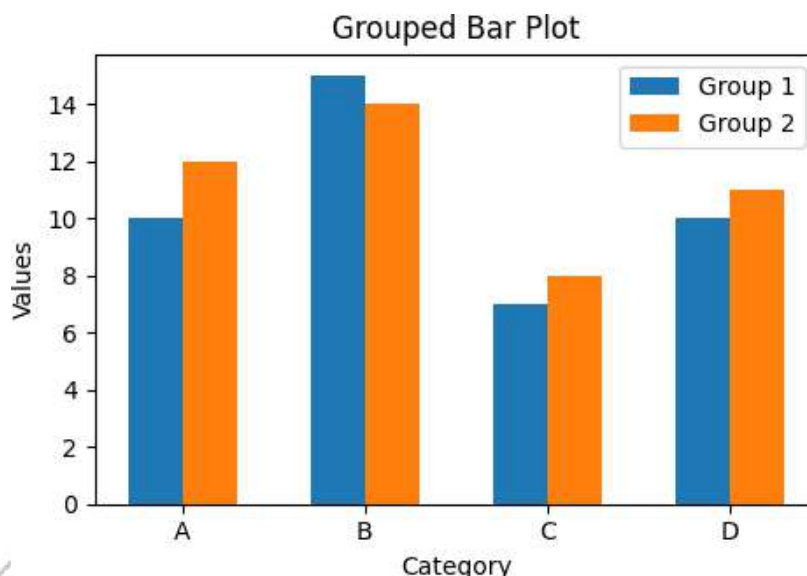
# Setting the positions and width for the bars
pos = np.arange(len(df['Category']))
width = 0.3

# Plotting the grouped bars
fig, ax = plt.subplots()
bars1 = ax.bar(pos - width/2, df['Group1'], width, label='Group 1')
bars2 = ax.bar(pos + width/2, df['Group2'], width, label='Group 2')

# Adding labels, title, and custom x-axis tick labels
ax.set_xlabel('Category')
ax.set_ylabel('Values')
ax.set_title('Grouped Bar Plot')
ax.set_xticks(pos)
ax.set_xticklabels(df['Category'])
ax.legend()

# Displaying the plot
plt.tight_layout()
plt.show()
```

Output:



Explanation of code:

- Import Libraries:** We import necessary libraries pandas for data handling and matplotlib.pyplot for plotting.
- Sample Dataset:** We create a sample dataset (data) which includes categories (Category) and two groups (Group1 and Group2).
- DataFrame Creation:** We convert the data dictionary into a Pandas DataFrame df.
- Setting Positions and Width:** We use numpy to create positions (pos) and set the width (width) for the grouped bars.
- Plotting:** We create a figure and axis (fig, ax) using Matplotlib. Then, we plot two sets of bars (bars1 and bars2) representing Group1 and Group2 respectively, using the positions and widths defined.
- Labels and Legends:** We set labels for axes, a title for the plot, custom x-axis tick labels (Category values), and a legend to differentiate between Group1 and Group2.
- Displaying the Plot:** Finally, we display the plot using plt.show().

You can modify the data dictionary to load your dataset from a CSV file using `pd.read_csv('filename.csv')`. This example provides a basic framework for plotting grouped bars in Python using Pandas and Matplotlib.

IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

- ## XI Conclusion

XII Practical Related Questions

1. What is grouped bars?
2. How to load dataset using `.csv` file?
3. How to plot a bar chart with multiple labels in Matplotlib?

[Space for Answer]

XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.projectpro.io/recipes/generate-grouped-bar-plot-in-python>
2. <https://www.geeksforgeeks.org/create-a-grouped-bar-plot-in-matplotlib>
3. <https://www.tutorialspoint.com/create-a-grouped-bar-plot-in-matplotlib>
4. <https://www.geeksforgeeks.org/plot-data-from-excel-file-in-matplotlib-python/>
5. How to plot a Bar Chart with multiple labels in Matplotlib (tutorialspoint.com)

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	

Practical No.15: Implementation of a python program that loads any dataset and plot a pie chart.

I Practical Significance

Implementing a Python program that loads any dataset and plots a pie chart is immensely valuable for data visualization and analysis. Pie charts are a popular tool for representing categorical data, providing a clear visual distinction of how different parts contribute to the whole. This program allows users to quickly and effectively transform raw data into an easily interpretable visual format, aiding in decision-making processes and presentations. Whether it's for business analytics, academic research, or personal projects, being able to generate pie charts from datasets helps in summarizing and communicating complex data insights effectively.

II Industry/ Employer Expected Outcome

1. To develop proficiency in creating impactful charts, graphs, and dashboards.
2. To use of tools like Tableau, Power BI, or Python for visualization.

III Course Level Learning Outcome

CO5-Create data visualization using many distributions.

IV Laboratory Learning Outcome

- LLO 15.1-Perform dataset loading to plot a pie chart.
LLO 15.2-Write a Python program to create pie chart from the loaded data.

V Relevant Affective domain related Outcome(s)

1. Follow safety practices.
2. Maintain tools and equipment.
3. Follow ethical practices.

VI Relevant Theoretical Background

To implement a Python program for loading datasets and plotting pie charts, one needs a basic understanding of Python programming, including knowledge of libraries such as **pandas** for data manipulation and **matplotlib** for plotting. Familiarity with data structures like **DataFrames** and basic operations such as reading data from files (e.g., CSV, Excel) is essential. Understanding the concept of categorical data and how pie charts represent proportions is also important. Matplotlib's pie chart function (**plt.pie**) requires an understanding of how to pass data and configure chart properties to create meaningful visualizations.

VII Exercises

- Implement python program create pie chart for data visualization for any dataset

VIII Procedure/Steps to plot pie chart in Python:

Here's a sample Python program that loads a dataset and plots a pie chart. This example uses

the **pandas** library to handle the dataset and **matplotlib** to create the pie chart. We'll assume the dataset is in a CSV file and contains categorical data suitable for a pie chart.

Sample Code:

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
file_path = 'your_dataset.csv' # Replace with your dataset file
pathdata = pd.read_csv(file_path)

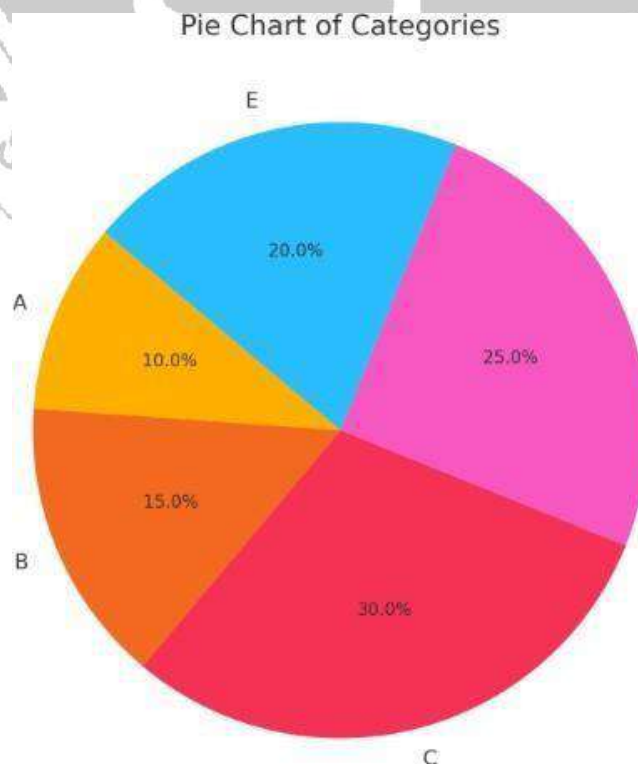
# Display the first few rows of the dataset to understand its structure
print(data.head())

# Assume the dataset has a column named 'Category' and another column named 'Value'
# You can adjust these column names based on your actual dataset
category_column = 'Category' # Replace with the name of the column containing categories
value_column = 'Value' # Replace with the name of the column containing values

# Aggregate the data if necessary (e.g., summing values for each category)
aggregated_data = data.groupby(category_column)[value_column].sum()

# Plot the pie chart
plt.figure(figsize=(8, 8))
plt.pie(aggregated_data, labels=aggregated_data.index, autopct='%1.1f%%', startangle=140)
plt.title('Pie Chart of Categories')
plt.show()
```

Output:



Explanation of code:

1. **Import Libraries:** We import **pandas** for data handling and **matplotlib.pyplot** for plotting.
2. **Load the Dataset:** The **pd.read_csv** function loads the dataset from a CSV file.
3. **Inspect the Dataset:** The **print (data.head())** statement displays the first few rows of the dataset to help understand its structure.
4. **Specify Columns:** We assume there are two relevant columns: one for categories and one for their corresponding values.
5. **Aggregate Data:** If necessary, we aggregate the data by summing the values for each category using the **groupby** function.
6. **Plot the Pie Chart:** We use **plt.pie** to plot the pie chart, setting labels to the category names and displaying the percentage of each category.

Example CSV Dataset (your_dataset.csv):

Category, Value
A, 10
B, 15
C, 30
D, 25
E, 20

Replace '**your_dataset.csv**' with the path to your actual dataset file and adjust the column names accordingly to match your data. This program will read the dataset, process the data, and display a pie chart representing the distribution of categories.

IX Required Resources

Sr. No.	Name of Resource	Major Specification	Qty.	Remarks
1.	Computer System	Any desktop or laptop computer with CPU> i3 and RAM 4GB onwards	One computer system for each student	
2.	Software Package	Python IDE, Video makers/editors, Visualization tools		

X Precautions to be Followed

1. Double-check your data for accuracy before presenting it.
2. Explain technical terms if necessary.
3. Provide context for the data to help the audience understand its significance.

XI Conclusion

.....

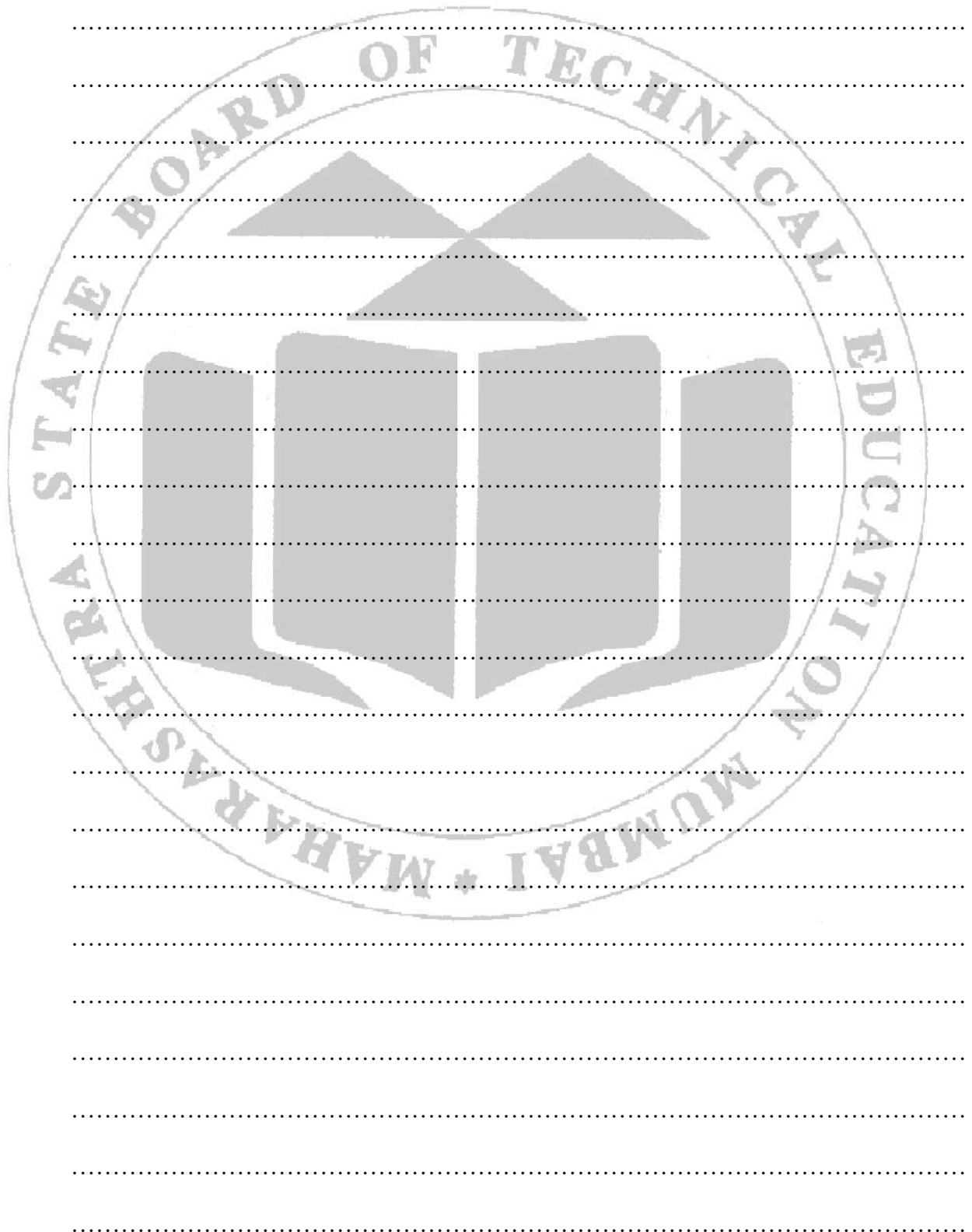
.....

XII Practical Related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. What the different type of charts?
2. How to plot a pie chart with multiple labels in Matplotlib?

[Space for Answer]



XIII References / Suggestions for further Reading Software/Learning Websites

1. <https://www.pythoncharts.com/matplotlib/pie-chart-matplotlib/>
2. <https://www.geeksforgeeks.org/plot-a-pie-chart-in-python-using-matplotlib/>

XIV Assessment Scheme

Performance indicators		Weightage
Process related: 15 Marks		60%
1.	Debugging ability	20%
2.	Correctness of Program codes	30%
3.	Quality of output achieved (LLO mapped)	10%
Product related: 10 Marks		40%
1.	Completion and submission of practical in time	20%
2.	Answer to sample questions	20%
Total : 25 Marks		100%

Marks obtained			Dated Sign of Teacher
Process Related (15)	Product Related (10)	Total (25)	