



GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor/Master of Engineering Syllabus

Subject Code : 3175202

Subject Name : Exploratory Data Analysis

WEF Academic Year :	2021-22
Semester :	7
Category of the Course :	Professional Core

Prerequisite :	-
Rationale :	To learn the essential exploratory techniques for analyzing and visualizing data, and to gain hands-on experience of using software tools for data analytics.

Course Scheme :

Teaching Scheme			Total Credits	Assessment Pattern and Marks				Total Marks
L	T	PR		C	Theory		Practical	
			ESE (E)		PA(M)	ESE (V)	PA (I)	
3	0	2	4	70	30	30	20	150

Course Content :

Sr. No.	Course Content	No. of Hours	% of Weightage
1	INTRODUCTION TO EXPLORATORY DATA ANALYSIS AND DISTRIBUTIONS: Basic Concepts—Population and Sample, A Case Study: Visually Inspecting Data to Improve Product Quality, Pareto Diagrams and Dot Diagrams, Frequency Distributions, Graphs of Frequency Distributions, Stem-and-Leaf Displays, Descriptive Measures, Quartiles and Percentiles, calculation of \bar{X} and S, Problems with aggregating data, Sample Spaces and Events, Counting, Probability, The Axioms of Probability, Some Elementary Theorems, Conditional Probability, Bayes' Theorem.	10	24
2	PROBABILITY MASS FUNCTION, CUMULATIVE DISTRIBUTIONS AND MODELING DISTRIBUTIONS: Making a Basic Histogram, Making Multiple Histograms from Grouped Data, Making a Density Curve, Making Multiple Density Curves from Grouped Data, Making a Frequency Polygon, Making a Basic Box Plot, Adding Notches to a Box Plot, Adding Means to a Box Plot, Making a Violin Plot, Making a Dot Plot, Making Multiple Dot Plots for Grouped Data, Making a Density Plot of Two-Dimensional Data, Miscellaneous Graphs.	10	24
3	RELATIONSHIP BETWEEN VARIABLES AND ESTIMATION: Scatter Plots, Characterizing Relationships, Correlation, Covariance, Pearson's		



GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor/Master of Engineering Syllabus

Subject Code : 3175202

Subject Name : Exploratory Data Analysis

	Correlation, Nonlinear Relationships, Spearman's Rank Correlation, Correlation and Causation, The Estimation Game, Guess the Variance, Sampling Distributions, Sampling Bias, Exponential Distributions, Classical Hypothesis Testing, Hypothesis Test, Testing a Difference in Means, Other Test Statistics, Testing a Correlation, Testing Proportions, Chi-Squared Tests.	12	28
4	TIME SERIES AND SURVIVAL ANALYSIS: Survival Curves, Hazard Function, Estimating Survival Curves, Kaplan-Meier Estimation, The Marriage Curve, Estimating the Survival Function, Confidence Intervals, Normal Distributions, Sampling Distributions, Representing Normal Distributions, Central Limit Theorem, Testing the CLT, Applying the CLT, Correlation Test, Chi-Squared Test.	10	24
	Total	42	100

Reference Book:

1. Think Stats, 2nd Edition: Exploratory Data Analysis, Allen B. Downey
2. Making sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining, by Glenn J. Myatt.
3. Making Sense of Data II: A Practical Guide to Data Visualization, Advanced Data Mining Methods, and Applications, Glenn J. Myatt, and Wayne P. Johnson.
4. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron
5. Online resources: Kaggle datasets, UCI Machine Learning Repository, etc.

Course Outcome:

After Completion of the Course, the Student will able to:

No	Course Outcomes	RBT Level*
01	Understand the philosophy of exploratory data analysis	UN
02	List various techniques for efficient visualization	RN
03	Apply discrete and continuous probability distributions and estimate the correlation between variables	AP
04	Analyze linear and nonlinear models visually	AN
05	Visualize time series and survival evaluation	EL
06	Use various visualization structures such as tables, spatial data, time-varying data, tree and network	AP

*RM: Remember, UN: Understand, AP: Apply, AN: Analyze, EL: Evaluate, CR: Create

Suggested Course Practical List:

Sample Case study based practical
Online Retail Sales Analysis



GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor/Master of Engineering Syllabus

Subject Code : 3175202

Subject Name : Exploratory Data Analysis

Scenario: You're a data analyst working for an e-commerce company. Your task is to perform exploratory data analysis on the company's sales data to uncover insights that could help improve business strategies.

- **Step 1: Data Understanding and Preparation**
Data Source: Obtain the dataset containing sales data, including information such as product details, sales quantities, prices, customer demographics, etc.
Data Cleaning:
Handle missing values: Check for missing values in important columns like sales quantity, prices, etc., and decide how to handle them (remove, impute).
Check for duplicates and outliers in the dataset.
- **Step 2: Summary Statistics and Data Visualization**
Summary Statistics:
Compute basic statistics for sales quantities, prices, and other relevant numerical variables.
Univariate Analysis:
Plot histograms or box plots to visualize the distribution of sales quantities, prices, etc.
Bivariate Analysis:
Explore relationships between sales quantities and prices, sales and customer demographics (age, gender, location), etc.
- **Step 3: Time-Series Analysis**
If the dataset includes timestamps, conduct time-series analysis to:
Identify trends and seasonality in sales data.
Analyze sales performance over different time periods (daily, weekly, monthly).
- **Step 4: Customer Segmentation**
Use clustering techniques to segment customers based on their purchasing behavior, such as:
RFM (Recency, Frequency, Monetary) analysis to classify customers into segments like high-value, low-value, etc.
Analyze purchasing patterns of different customer segments.
- **Step 5: Product Analysis**
Analyze product performance by:
Identifying top-selling products.
Analyzing product sales trends over time.
Investigating the relationship between product attributes (category, brand, etc.) and sales.
- **Step 6: Geographic Analysis**
Explore sales patterns across different geographic regions:
Plot geographical heatmaps to visualize sales distribution.
Analyze differences in sales performance between regions.
- **Step 7: Seasonal Trends and Promotions**
Analyze the impact of seasonal trends and promotions on sales:
Identify peak sales periods (holidays, special events).
Evaluate the effectiveness of past promotions on sales.
- **Step 8: Business Insights and Recommendations**
Summarize key findings and insights from the EDA process.
Provide actionable recommendations to improve business strategies, such as:
Targeting specific customer segments with personalized marketing campaigns.
Adjusting pricing or product offerings based on sales trends and customer preferences.



GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor/Master of Engineering Syllabus

Subject Code : 3175202

Subject Name : Exploratory Data Analysis

- **Step 9: Visualization and Reporting**
Create visually appealing charts and graphs to present key findings.
Prepare a comprehensive report or presentation summarizing the EDA results and recommendations for stakeholders.
- **Step 10: Iterative Analysis**
EDA is an iterative process. Revisit previous steps and refine analyses based on new insights or questions that arise during the analysis.

List of Laboratory/Learning Resources Required:

* * * * *