Statistics, Statistical Modelling & Data Analytics – Case Studies:

All students are required to submit the Case studies by 17 March 2025. Below are the list of Case Studies. A group has been formulated for this task. All should work in the group.

Note: Make a front page with the Title allotted, Name and Roll Number of all the students in the group. Then complete the case study with the questions given below. You need to mention the contribution of each student at the end of the Case Study.

Case Study No.	Case Study	Roll No.	Name
Case Study 1	Customer Segmentation using Descriptive Statistics and Probability	00214811122	MAYANK AGARWAL
		00314808223	PULKIT KUMAR
		00314811122	SHIVANSH NEGI
		00414811122	ANIRUDH NAIR
		00514808223	NAVNEET KUMAR SHARMA
Case Study 2	Predicting House Prices with Regression Models	00614808223	ASHISH KUMAR KASHYAP
		00614811122	AYUSH AHUJA
		00714808223	VASU KUMAR
		00814808223	VANSH CHOUDHARY
		00814811122	JAYANT
Case Study 3	Data Analytics in Healthcare - Diagnosing Diabetes	00914808223	HARSH RAJ SINGH
		01014808223	PARV YADAV
		01014811122	SARTHAK
		01114808223	PRIYANSHU DHARWAL
		01214808223	KESHAV NANDAN
Case Study 4	Eigenvalues and Eigenvectors in Image Compression	01314808223	MD TAUFIQUE UMAR
		01414811122	PRABHAT SHANKAR DUBEY
		01514811122	AGRIM SETH
		01614811122	RISHANK DABAS
		02214811122	PRERIT SHARMA
Case Study 5	Fraud Detection in Banking using Hypothesis Testing	02614811122	AAKRISHT SHARMA
		03214811122	ARPIT MISHRA
		03314811122	ATHARV SINGH
		03714811122	ISHAAN JHA
		03914811122	PANKAJ

Case Study 6	Analyzing Social Media Sentiments with Logistic Regression	04214811122	SANSKAR
		04314811122	AKSHIT DAS
		04614811122	PANKAJ KUMAR
		04714811122	DHRUV
		20114811122	VIBUDH YADAV
Case Study 7	Quality Control in Manufacturing using Sampling Distributions	35114811122	PRISHA SAGAR
		35314811122	NISHCHAY RANA
		70214811122	SHREYANSH PRATAP RAO
		70314811122	NAMAN BAGAI

Case Studies Based on University Syllabus

Case Study 1: Customer Segmentation using Descriptive Statistics and Probability

Topic: Statistics & Descriptive Analytics (Unit-I)
Scenario:
A retail company wants to understand its customer base to tailor its marketing campaigns. Using historical purchase data, the company applies descriptive statistical methods (mean, median, mode, variance) and probability distributions to identify key customer segments.
Key Questions:

- How can the company use measures of central tendency to group customers?
- What probability distribution best represents customer spending behavior?
- How can hypothesis testing help in validating marketing strategies?

Case Study 2: Predicting House Prices with Regression Models

Topic: Statistical Modelling & Regression Analysis (Unit-II)
Scenario:
A real estate firm aims to develop a predictive model for house prices based on variables such as location, size, and number of bedrooms. By applying linear regression and the Gauss-Markov theorem, they optimize their pricing strategy.
Key Questions:

- What independent variables have the highest impact on house prices?
- How does multicollinearity affect regression results?
- How can residual analysis improve model accuracy?

Case Study 3: Data Analytics in Healthcare - Diagnosing Diabetes

Topic: Metric Space & Compactness in Data Analytics (Unit-III)
Scenario:
A healthcare provider wants to classify patients as diabetic or non-diabetic based on medical test results. By using the concept of metric space and compactness, they optimize clustering techniques to detect patterns in patient data.
Key Questions:

- How can metric space concepts help in clustering patient data?
- What role does compactness play in classifying medical conditions?
- How can completeness and connectedness improve disease prediction models?

Case Study 4: Eigenvalues and Eigenvectors in Image Compression

Topic: Eigenvalues & Eigenvectors in Data Analytics (Unit-IV)
Scenario:
A tech company is working on an image compression algorithm using Principal Component Analysis (PCA). They utilize eigenvalues and eigenvectors to reduce image dimensions while preserving essential features.
Key Questions:

- How do eigenvalues and eigenvectors contribute to dimensionality reduction?
- What is the significance of basis and subspace in PCA?
- How can vector independence improve image compression efficiency?

Case Study 5: Fraud Detection in Banking using Hypothesis Testing

Topic: Hypothesis Testing & Statistical Inference (Unit-I)
Scenario:
A bank is facing increasing cases of fraudulent transactions and wants to identify anomalies in customer behavior. Using hypothesis testing and probability distributions, they develop a fraud detection system.
Key Questions:

- How can hypothesis testing be used to detect fraudulent transactions?
- What statistical inference techniques are useful for anomaly detection?
- How can probability distributions help in setting fraud detection thresholds?

Case Study 6: Analyzing Social Media Sentiments with Logistic Regression

Topic: Logistic Regression & Poisson Regression (Unit-II)
Scenario:
A marketing agency wants to predict whether a tweet is positive, negative, or neutral based on text analysis. They use logistic regression to classify sentiments and Poisson regression to analyze tweet frequency trends.
Key Questions:

- How does logistic regression help in binary classification problems?
- What role does Poisson regression play in analyzing tweet frequencies?
- How can model selection improve the accuracy of sentiment prediction?

Case Study 7: Quality Control in Manufacturing using Sampling Distributions

Topic: Sampling Distributions & Statistical Inference (Unit-I)
Scenario:
A manufacturing company wants to ensure that the products meet quality standards. They implement a statistical sampling approach to assess defect rates in batches.
Key Questions:

- How can sampling distributions help in quality control?
- What statistical inference techniques can be applied to improve manufacturing processes?
- How can variance analysis help in identifying process inefficiencies?
