Ramaiah Institute of Technology
(Autonomous Institute, Affiliated to VTU)

Data Analysis Using R (23CSAEC49)

Academic year 2025 - 2026

Project Based Learning (20 Marks)

By : Jamuna S Murthy

Suggested Titles

1. **Predictive Analytics on Stock Market Data Using Time Series Models**
   - Line plot of stock prices over time.
   - Moving averages and stock price volatility.
   - Correlation heatmap between different stock prices.
   - Box plot for stock returns by month.
2. **Exploring Customer Segmentation with K-means Clustering on Retail Data**
   - Scatter plot of customer features colored by cluster.
   - Elbow plot to determine the optimal number of clusters.
   - Pairplot of features used in clustering.
   - 3D scatter plot showing customer segmentation.
3. **Sentiment Analysis of Tweets Using Text Mining Techniques**
   - Word cloud of most frequent words in tweets.
   - Sentiment distribution bar plot (positive, negative, neutral).
   - Sentiment analysis over time (e.g., sentiment by date).
   - Pie chart of sentiment categories.
4. **Visualizing Climate Change Data and Predicting Future Trends**
   - Line chart showing global temperature trends over time.
   - Heatmap of temperature anomalies by region.
   - Bar plot comparing emissions by country.
   - Scatter plot of CO2 levels vs. global temperature.
5. **Developing a Recommendation System for E-commerce Using Collaborative Filtering**
   - Heatmap of user-item interaction matrix.
   - Distribution of item ratings across users.
   - Bar plot of the top recommended products.
   - 3D plot of user preferences based on the clustering of similar users.

6. **Customer Churn Prediction Using Logistic Regression and Random Forest**
    - ROC curve to evaluate model performance.
    - Feature importance bar chart from random forest.
    - Confusion matrix visualization.
    - Pie chart of churn vs non-churn customers.
7. **Analyzing Disease Outbreaks with Epidemiological Data**
    - Line graph showing the growth of cases over time.
    - Geospatial heatmap of case density by region.
    - Scatter plot of cases vs. vaccination rates.
    - Bar chart of total cases per region or country.
8. **Crime Data Analysis and Prediction with Geospatial Techniques**
    - Heatmap of crime hotspots by area.
    - Bar chart showing crime frequency by category.
    - Time series plot of crime rates over the years.
    - Map visualization of crime data locations.
9. **Predicting House Prices Using Multiple Regression and Feature Engineering**
    - Scatter plot of house prices vs. square footage or other key features.
    - Correlation matrix heatmap of features.
    - Box plot of house prices by neighborhood.
    - Residual plot to assess model fit.
10. **Market Basket Analysis for Product Association Using Apriori Algorithm**
    - Bar plot of frequent itemsets.
    - Network graph of product associations.
    - Heatmap showing support, confidence, and lift for item pairs.
    - Sankey diagram of product flow through transactions.
11. **Analysis of COVID-19 Data to Identify Key Factors in Disease Spread**
    - Line plot of new cases over time.
    - Map of confirmed cases by country or state.
    - Histogram of age group distribution among cases.
    - Box plot comparing recovery times by region.
12. **Sentiment Analysis of Movie Reviews Using Natural Language Processing (NLP)**
    - Word cloud of most frequent words in reviews.
    - Sentiment distribution plot for each movie.
    - Time series plot showing sentiment trend over time.
    - Bar plot of movie ratings by sentiment.
13. **Stock Portfolio Optimization Using Modern Portfolio Theory (MPT)**
    - Efficient frontier plot showing risk vs. return.
    - Bar plot comparing portfolio returns with individual stocks.

- Scatter plot showing correlation between stocks in the portfolio.
- Pie chart of portfolio asset allocation.

14. **Time Series Forecasting of Air Traffic using ARIMA Models**
    - Line plot of historical air traffic data.
    - Forecast plot showing actual vs predicted values.
    - Autocorrelation plot (ACF/PACF).
    - Decomposition of time series into trend, seasonality, and residuals.
15. **Prediction of Loan Default Using Machine Learning Techniques**
    - ROC curve for model evaluation.
    - Feature importance plot.
    - Confusion matrix showing prediction accuracy.
    - Histogram of loan amounts for default vs non-default customers.
16. **Understanding Customer Preferences with Survey Data Analysis**
    - Bar chart showing responses to survey questions.
    - Pie chart of demographic distribution of respondents.
    - Heatmap of responses to multiple survey questions.
    - Box plot of satisfaction scores by customer segment.
17. **Visualizing and Predicting Sales Trends in Retail Using R**
    - Line graph of historical sales trends.
    - Forecasting plot using ARIMA or exponential smoothing.
    - Seasonal decomposition of sales data.
    - Bar plot comparing sales by region or product.
18. **Predicting Election Results Based on Historical Data Using Classification Models**
    - Bar chart showing historical election results by region.
    - Confusion matrix for model classification.
    - Pie chart of party distribution in an election.
    - Line chart showing poll results vs. actual results over time.
19. **Clustering of Species Based on Environmental Data Using Hierarchical Clustering**
    - Dendrogram showing hierarchical clustering.
    - Scatter plot of species distribution by cluster.
    - Heatmap of environmental factors influencing clustering.
    - Box plot of species distribution by cluster.
20. **Analysis of Education Data to Predict Student Performance**
    - Scatter plot of study hours vs. student grades.
    - Box plot of grades by gender or other demographics.
    - Heatmap showing correlations between student features.
    - Histogram of student performance distribution.
21. **Social Media Engagement Analysis Using User Activity Data**

- Bar plot of engagement frequency by time of day.
- Line chart of follower growth over time.
- Sentiment analysis bar chart of social media posts.
- Word cloud of most popular hashtags.

22. **Predictive Modeling for Healthcare Cost Estimation Using Regression Analysis**
    - Line plot of healthcare costs over time.
    - Scatter plot of age vs. healthcare expenditure.
    - Feature importance bar chart.
    - Histogram of cost distribution by insurance type.
23. **Sentiment Analysis of Product Reviews and Customer Feedback**
    - Word cloud of product review keywords.
    - Pie chart of sentiment distribution (positive, negative, neutral).
    - Time series plot of sentiment over reviews.
    - Bar chart of average ratings by product category.
24. **Air Quality Index Prediction Using Machine Learning**
    - Line plot showing air quality index trends over time.
    - Heatmap of air quality by region.
    - Scatter plot of pollutants vs. air quality index.
    - Seasonal decomposition of air quality index.
25. **Classifying Email Messages as Spam or Not Spam Using NLP Techniques**
    - Word cloud of frequent words in spam vs. non-spam emails.
    - Confusion matrix showing model accuracy.
    - ROC curve for spam detection model.
    - Bar chart showing frequency of common spam keywords.
26. **Exploratory Data Analysis of Global Tourism Trends Using Open Datasets**
    - Line graph showing global tourism trends by year.
    - Bar plot comparing tourism numbers by country.
    - Heatmap of top tourist destinations by region.
    - Scatter plot of tourism expenditure vs. visitor count.
27. **Identifying Factors Affecting Employee Retention Using HR Data**
    - Bar plot of retention rate by department.
    - Scatter plot of salary vs. employee tenure.
    - Heatmap showing correlations between retention factors.
    - Box plot comparing retention by demographic group.
28. **Weather Forecasting Using Historical Weather Data**
    - Line chart of temperature over time.
    - Box plot of temperature variation by month.
    - Scatter plot of humidity vs. temperature.
    - Time series forecast of temperature for the next month.

29. **Analysis of Global Economic Indicators and Their Impact on National Growth**
    - Line plot of GDP growth vs. key economic indicators.
    - Scatter plot of inflation vs. unemployment rate.
    - Heatmap of correlation between economic factors.
    - Bar chart comparing economic performance of countries.
30. **Building a Time Series Forecasting Model for Energy Consumption Prediction**
    - Line plot of energy consumption over time.
    - Decomposition of energy consumption into trend, seasonality, and noise.
    - Forecasting plot of energy consumption.
    - Heatmap of energy consumption by region and time of day.