Retail & Marketing Analytics Project For

Customer Segmentation & Growth Strategy

A Comprehensive End to End Analytics Project for Students

Duration: 2-3 Hours | Recorded Lectures Available

@ashoks773

@ashoks773

@sharma-ak

Instructor: Dr. Ashok Kumar Sharma
Bioinformatics Scientist (Senior Data Analyst)
Los Angeles, CA

Project links:
https://www.skillsetmaster.com/data-analytics

Tools: Python | Pandas | Numpy | Scikit-learn | Scipy | Statsmodels | Matplotlib | Seaborn | Plotly

Target Roles: Marketing Analyst; Business Intelligence Analyst; Data Analyst; Retail Analytics Specialist

Follow: @datawithashok YouTube

About Dataset

Dataset:

- **Primary**: Kaggle - Retail Sales Dataset
- **Type**: Transactional retail data
- **Size**: ~10,000 transactions
- **Time Period**: 2022-2024 (adjustable based on actual data)

Key Features:

Feature	Description	Type
Order_ID	Unique order identifier	String
Order_Date	Date of purchase	DateTime
Customer_ID	Unique customer identifier	String
Product_ID	Unique product identifier	String
Product_Category	Product category	Categorical
Sales	Order revenue	Numeric
Quantity	Items purchased	Numeric
Discount	Discount applied	Numeric
Region	Geographic region	Categorical
Segment	Customer segment	Categorical

Technologies Used:

Programming & Analysis

- **Python 3.9+**: Core programming language
- **pandas**: Data manipulation and analysis
- **NumPy**: Numerical computing
- **scikit-learn**: Machine learning (K-Means cluste
- **scipy**: Statistical analysis

Visualization

- **Matplotlib**: Static visualizations
- **Seaborn**: Statistical graphics
- **Plotly**: Interactive charts
- **Power BI / Tableau**: Business intelligence dashboards

Development Tools

- **Jupyter Notebook**: Interactive development
- **Git & GitHub**: Version control
- **VS Code**: Code editor

Project Overview – Objectives

This comprehensive analytics project demonstrates a complete data science workflow for retail and marketing analytics, from data acquisition to actionable insights and dashboard creation.

Objectives

1. **Customer Segmentation**: Identify distinct customer groups using RFM analysis and K-Means clustering
2. **Sales Analysis**: Analyze sales trends, seasonality, and product performance
3. **Marketing Optimization**: Evaluate customer lifetime value and marketing efficiency
4. **KPI Design**: Create comprehensive metrics framework for business monitoring
5. **Visualization**: Build interactive dashboards for stakeholder communication

Project Overview – Business related

Business Problem

A multi-channel retail company faces several challenges:

- **Declining customer retention** rates (42% vs industry benchmark of 45-50%)
- **Limited understanding** of customer behavior and preferences
- **Inefficient marketing** spend without proper segmentation
- **Inconsistent sales** performance across product categories and regions
- **Need for data-driven** decision making

Business Questions

1. Who are our most valuable customers?
2. What drives customer retention and churn?
3. How can we optimize our product mix?
4. Which marketing strategies deliver the best ROI?
5. What are the key performance indicators we should track?

Project Workflow

Phase 1: Data Preparation (Refer to Notebook)

- Data acquisition from Kaggle
- Initial data quality assessment
- Missing value treatment
- Outlier detection and handling
- Feature engineering
- Data type optimization

Phase 2: Exploratory Analysis (Refer to Notebook)

- Univariate analysis (distributions)
- Bivariate analysis (relationships)
- Time series patterns
- Customer behavior analysis
- Product performance analysis
- Correlation analysis

Phase 3: Advanced Analytics (Refer to Notebook)

RFM Analysis (Recency, Frequency, Monetary)
K-Means Clustering (4 customer segments)
Cluster profiling and naming
Market Basket Analysis (optional)
Cohort retention analysis
Customer Lifetime Value calculation

Phase 4: KPI & Reporting (Refer to Notebook)

- Comprehensive KPI framework
- Monthly trend analysis
- Category and regional performance
- Executive summary report
- Dashboard data preparation
- Strategic recommendations

📋 END-TO-END

Learning Outcomes

Technical Skills

Data acquisition and ETL processes
Data cleaning and preprocessing techniques
Exploratory data analysis (EDA)
Statistical analysis and hypothesis testing
Machine learning (K-Means clustering)
Customer segmentation (RFM analysis)
Data visualization best practices
Dashboard design and creation
SQL and database concepts
Python programming for data analysis

Business Skills

Retail and marketing analytics
KPI design and measurement
Customer lifetime value (CLV) calculation
Market basket analysis
Cohort analysis and retention metrics
Business insight generation
Executive communication
Strategic recommendation development

Tools & Technologies

Python (pandas, scikit-learn, matplotlib)
Jupyter Notebooks
Power BI / Tableau
Git & GitHub
Kaggle API

📁 Project Structure

```
retail-marketing-analytics/
│
├── data/
│   ├── raw/                          # Original datasets
│   │   └── retail_sales_data.csv
│   └── processed/                    # Cleaned and transformed data
│       ├── cleaned_retail_sales.csv
│       ├── rfm_analysis.csv
│       ├── customer_segments.csv
│       ├── customer_clv.csv
│       └── monthly_kpis.csv
│
├── notebooks/
│   ├── 01_Data_Acquisition_and_Setup.ipynb
│   ├── 02_Data_Cleaning_and_Preprocessing.ipynb
│   ├── 03_Exploratory_Data_Analysis.ipynb
│   ├── 04_Customer_Segmentation_and_Advanced_Analytics.ipynb
│   └── 05_KPI_Design_and_Dashboard_Preparation.ipynb
│
├── scripts/
│   ├── data_processing.py            # Data cleaning functions
│   ├── clustering.py                 # Segmentation algorithms
│   └── kpi_calculation.py            # KPI computation
│
├── dashboards/
│   ├── power_bi_dashboard.pbix       # Power BI dashboard file
│   └── README.md                     # Dashboard documentation
│
├── outputs/
│   ├── figures/                      # All visualizations (27+ charts)
│   │   ├── 01_missing_values.png
│   │   ├── 17_rfm_segments.html
│   │   └── ...
│   └── reports/                      # Analysis reports
│       ├── kpi_summary.csv
│       ├── executive_summary.txt
│       └── ...
│
├── docs/
│   ├── data_dictionary.csv           # Feature descriptions
│   ├── technical_documentation.md    # Technical details
│   └── presentation_slides.pdf       # Project presentation
│
├── .gitignore                        # Git ignore rules
├── requirements.txt                  # Python dependencies
├── README.md                         # This file
└── LICENSE                           # MIT License
```

Create GitHub Repo for the Project

```
# Add all files
git add .

# Create initial commit
git commit -m "Initial commit: Complete retail & marketing analytics project"

# Create GitHub repository (via GitHub website)
# Then connect local to remote
git remote add origin https://github.com/yourusername/retail-marketing-analytics.git

# Push to GitHub
git branch -M main
git push -u origin main
```