

Data Analysis Project

Project Title: Black Friday Sales Data Analysis
Data Analyst: Shijin Ramesh
Dataset Source: Kaggle – Black Friday Sales Dataset

Purpose:

To analyze customer purchasing behavior during the Black Friday sales event by exploring a large retail dataset. The project aims to uncover trends and relationships between demographic attributes (such as gender, age, occupation, and marital status) and purchasing patterns. The insights generated will be useful for understanding consumer segments and optimizing marketing strategies.

Scope / Major Project Activities:

Activity	Description
Data Collection	Load and inspect the dataset for structure, completeness, and validity
Data Cleaning	Handle missing values and remove or transform irrelevant columns
Column-wise Analysis	Analyze key columns like Gender, Age, Marital Status, and Product Category
Visualization	Generate charts (bar, pie, box plots) using Seaborn and Matplotlib
Multi-Column Relationships	Combine demographic factors to uncover layered purchasing patterns
Occupation & Product Insights	Examine relationships between occupation and product purchases
Grouped Statistical Analysis	Use groupby and aggregation to study purchases across different segments
Reporting	Summarize findings and prepare visual and written reports

This project does not include:

- Building recommendation models or forecasting algorithms
- Transactional pricing analysis or real-time behavioral tracking
- Predictive modeling using machine learning

Deliverables:

Deliverable	Description/Details
Cleaned Dataset	Preprocessed and cleaned version of the dataset
Jupyter Notebook (.ipynb)	Python code notebook with visualizations, explanations, and insights
Project Report (.pdf)	Executive summary with findings and key charts
README.md File	Overview of the project for GitHub presentation
requirements.txt	Python & R packages required to run the notebook

Schedule Overview / Major Milestones:

Milestone	Expected Completion Date	Description/Details
Initial Data Inspection	Day 1	Dataset loaded, structure checked, nulls identified
Column & Demographic Analysis	Day 2	Insights generated for Gender, Age, Marital Status etc.
Multi-Column & Grouped Insights	Day 3	Combined demographic segments analyzed
Final Report & GitHub Upload	Day 4	Notebook, report, and summary published

Estimated Completion Date:

Within 4 days of project start, assuming daily progress.