1. Summary of overall learning outcome of the internship

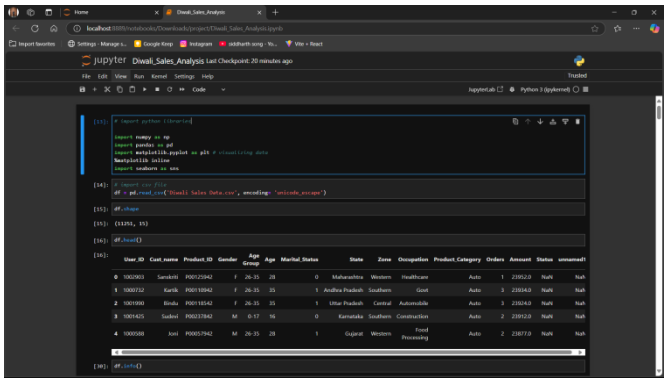
The Diwali Sales Analysis project aims to understand customer behavior during the Diwali season by analyzing real sales data. Using Python's powerful libraries like Pandas, NumPy, Matplotlib, and Seaborn, the project covers:

- Data cleaning
- Data transformation
- Exploratory Data Analysis (EDA)
- Visual storytelling through graphs and charts.

The goal is to find insights about who buys the most, what sells the most, and which groups have the highest purchasing power during Diwali.

.1.1Step-by-Step Code Explanation

- Importing Python Libraries



- Loading the Data

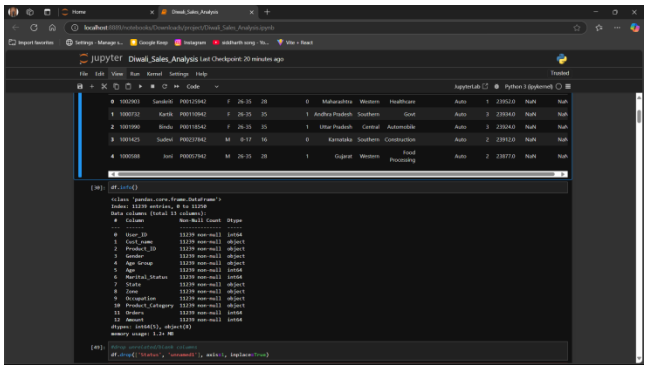
- Reading CSV: Loads the Diwali sales data into a DataFrame called df.
- Encoding: Some special characters in names/states need 'unicode_escape' encoding to avoid errors.
- df.shape: Tells us the number of rows and columns (11251 rows × 15 columns). Purpose: Bring raw data into Python for analysis.

- Initial Data Inspection

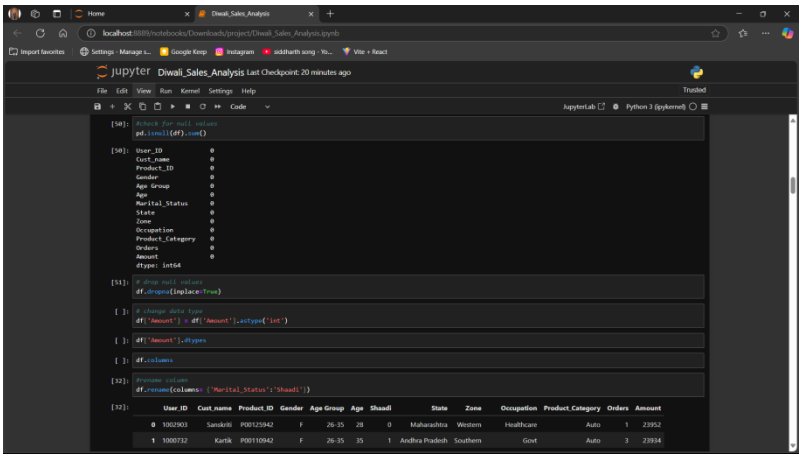
- head(): Displays the first 5 rows to quickly inspect the structure and content.
- info(): Provides the datatypes and null value counts for each column.

Findings:

- 15 columns exist, but 2 (Status, unnamed1) are blank or unused.
- Amount column has some null values.



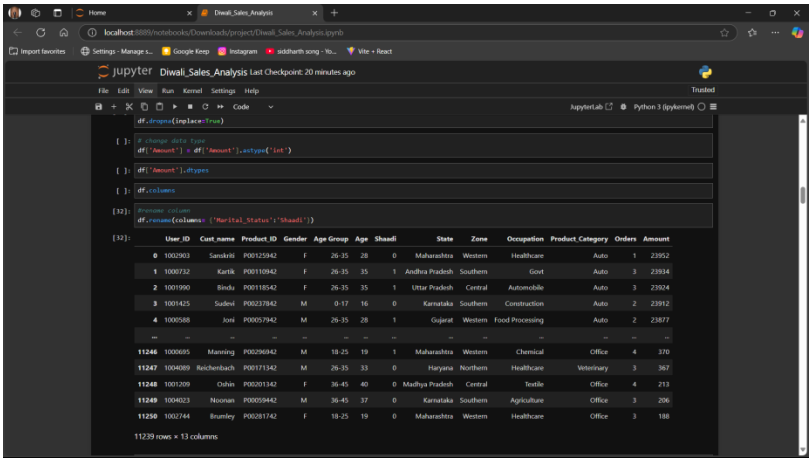
- Cleaning the Data



- Removes the irrelevant columns to clean up the dataset.
- Checks for missing values.
- Drops any rows where Amount is missing.

- Renaming Columns

  - Marital_Status is renamed internally for better readability.



- Descriptive Statistics

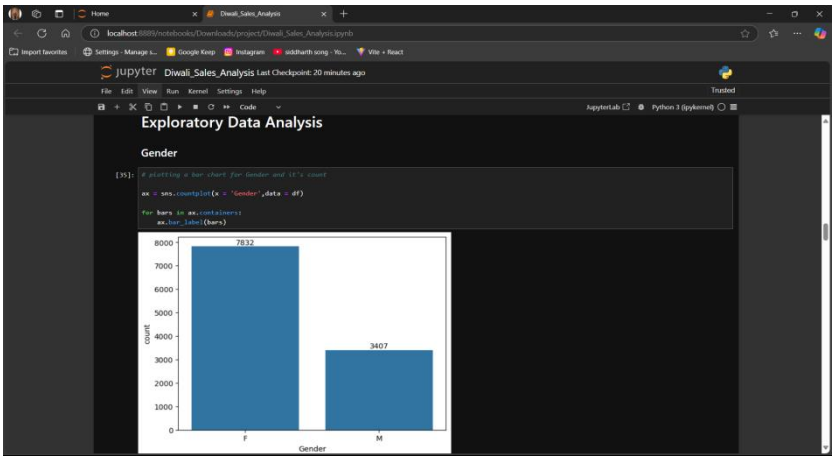
- Calculates mean, standard deviation, minimum, maximum, and quartiles.
- Focuses on numeric columns like Age, Orders, and Amount.

1.2 Exploratory Data Analysis (EDA)

- Gender Analysis

  - Gender Count Plot
  - Creates a bar chart showing the count of male and female customers.
  - Shows which gender spent more money.

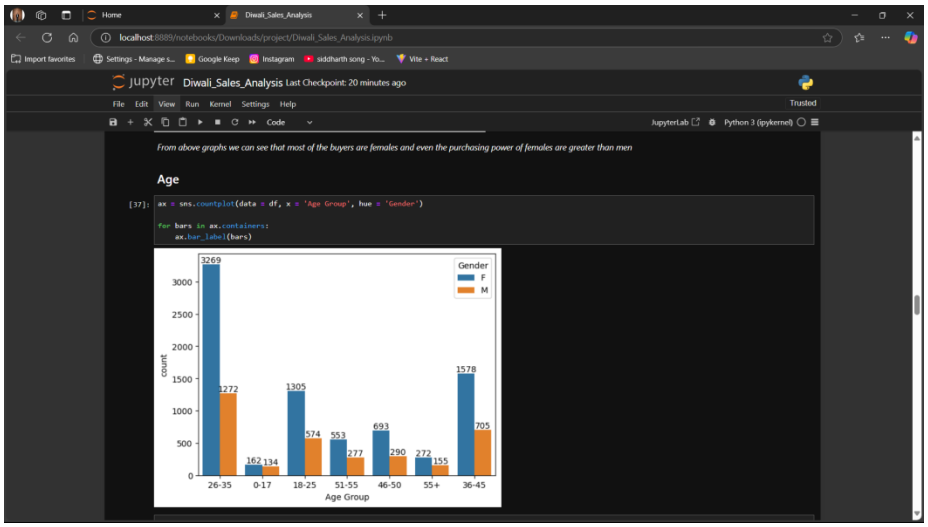
.



- Gender vs. Total Amount

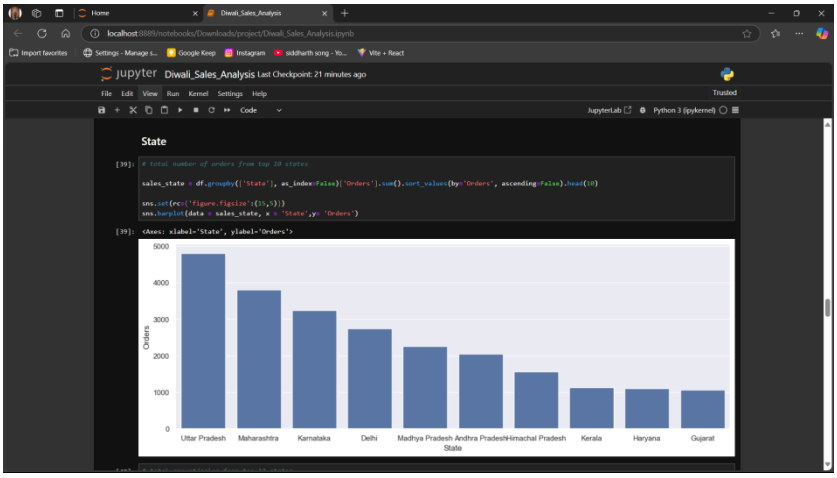
- Shows which gender spent more money.

- Females made more purchases and spent more overall compared to males



- State-wise Analysis

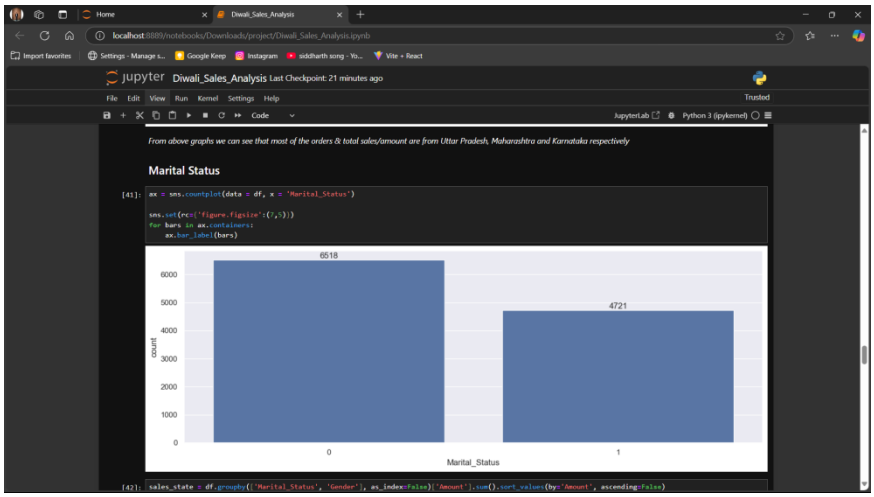
- Top States by Orders
- Top States by Total Amount
    - Insights:
- Uttar Pradesh, Maharashtra, and Karnataka are the top contributors in terms of both orders and revenue.



Marital Status Analysis

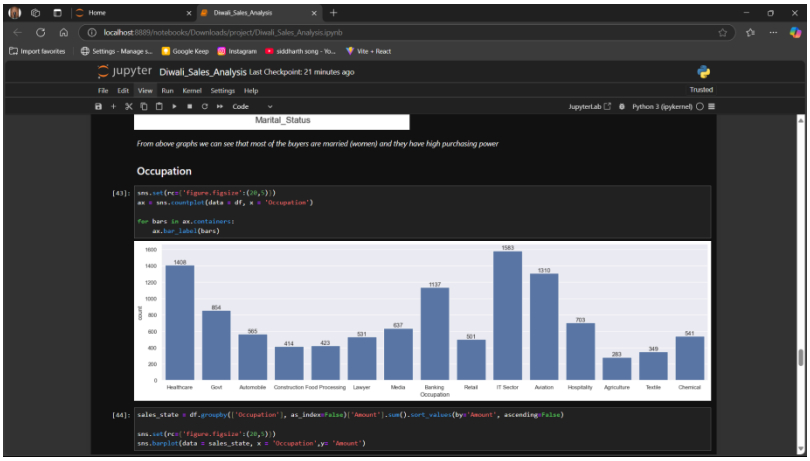
- Marital Status Count Plot
- Marital Status vs. Amount
- Insights:
- Married women spent the most.

- Marital status greatly impacts purchasing behavior.



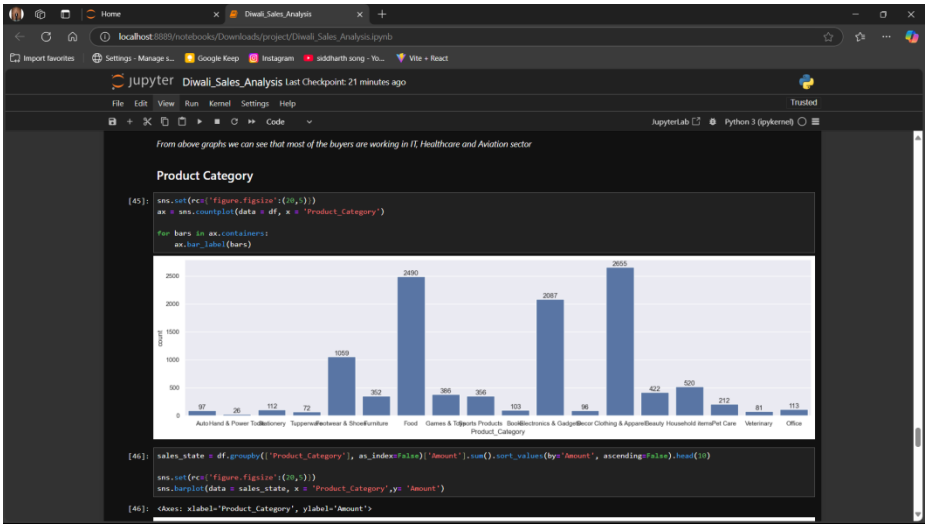
Occupation Analysis

- People working in IT, Healthcare, and Aviation sectors spend the most during Diwali.



Product Category Analysis

- Product Category Count
- Product Category by Sales Amount



Product Category Count Plot

- Purpose: Understand what categories customers are buying.
- Observation:
  Food, Clothing, and Electronics are the most popular.

Product Category vs. Amount Plot

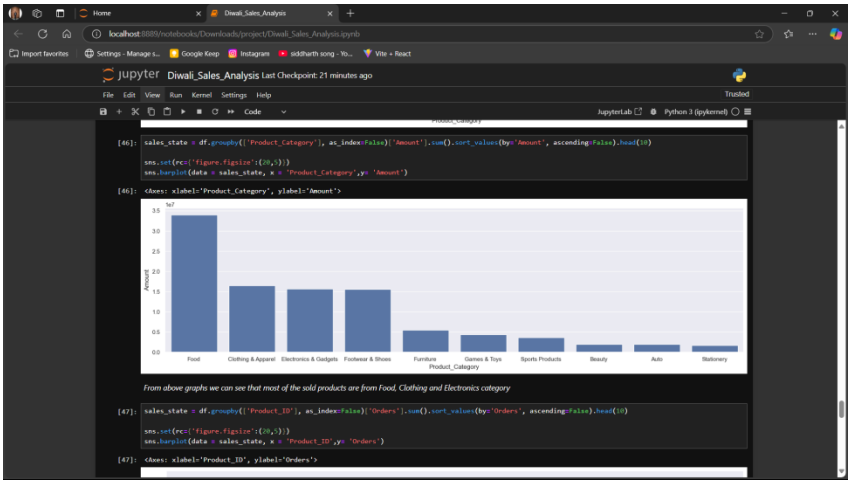
- Purpose: Understand revenue generation per category.
- Observation:
  Electronics generate a huge share of total revenue.

Business Application:

- Increase inventory and promotions on Food, Clothing, and Electronics categories during Diwali.
- Bundle offers (e.g., "Buy Food + Electronics and get a discount").

Possible Improvements:

- Study profitability per product category (not just revenue).



1.3 Top Selling Products (Expanded)

- Purpose: Identify which specific products are bestsellers.
- Observation:
  Certain product IDs have extremely high order counts.
- Business Application:
  - Stock more of these top-selling products.
  - Highlight them in advertisements and online platforms.
- Possible Improvements:
  - Find cross-sell patterns (what people often buy together).
  - Study return rates for top-selling products.