

Vending Machine Data Administrator – Project Report

Date: 24-10-2025

Prepared By: Medidha Sivaji

Dataset Used: VendingMachine_Refill_Operations_10k.csv

1. Executive Summary

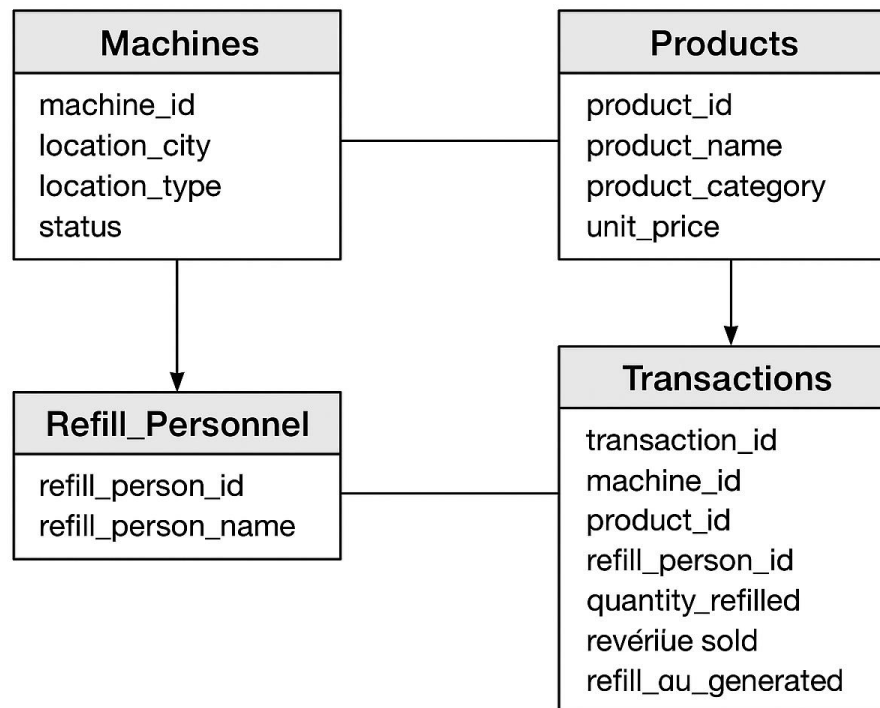
This report documents the complete process of designing a database, cleaning and analyzing vending machine operational data, and developing a Power BI dashboard to visualize key performance metrics. The objective was to evaluate data handling, analysis, and dashboard development capabilities using real-world vending machine refill and sales data. It covers database schema design, anomaly detection and cleaning, exploratory analysis, and dashboard creation.

2. Project Overview & Objectives

- Design a normalized relational database schema to efficiently store and query operational data.
- Clean the dataset and handle anomalies (missing, duplicate, or invalid entries).
- Perform exploratory data analysis to uncover patterns in refills, sales, and machine utilization.
- Develop an interactive Power BI dashboard showing KPIs such as revenue trends, refill frequency, and machine uptime.
- Document all findings, methodologies, and visual insights professionally.

3. Database Setup & Schema Design

The first step involved creating a normalized relational database to efficiently manage and query vending machine operations data. The dataset was imported into MySQL, and tables were designed following the principles of 3NF normalization.



Example SQL Schema:

```

CREATE TABLE Machines (
    machine_id VARCHAR(20) PRIMARY KEY,
    location_city VARCHAR(50),
    location_type VARCHAR(30),
    status VARCHAR(20)
);

CREATE TABLE Transactions (
    transaction_id VARCHAR(30) PRIMARY KEY,
    machine_id VARCHAR(20) REFERENCES Machines(machine_id),
    product_id VARCHAR(20),
    quantity_sold INT,
    revenue_generated DECIMAL(10,2),
    refill_date DATE
);
  
```

Data import was handled using MySQL Workbench's data import wizard. Basic data validation queries were executed to check record consistency and data integrity.

4. Data Cleaning & Preprocessing

Data cleaning was conducted using Python (Pandas and NumPy). Missing values, duplicates, and negative quantities were identified and corrected.

Example Python Code:

python

```
df = pd.read_csv("VendingMachine_Refill_Operations 1.csv") # Load dataset
df.dropna(subset=['revenue_generated'], inplace=True)
df = df[df['quantity_sold'] >= 0]
nan_counts = df[numeric_cols].isna().sum() # Check for NaN values in the numeric columns
print(nan_counts)
# Fill NaN values in numeric columns with 0
df[numeric_cols] = df[numeric_cols].fillna(0)
nan_counts = df[numeric_cols].isna().sum()
print(nan_counts) # Verify no NaN values remain

duplicates = df.duplicated(numeric_cols)
print(duplicates.sum()) # Print number of duplicate rows based on numeric columns
```

5.Exploratory Data Analysis & Insights

Exploratory Data Analysis (EDA) was performed to understand sales trends, refill frequencies, product performance, and operational efficiency. Insights were visualized using Power BI and Python (Matplotlib).

Key metrics analyzed:

- Monthly revenue trend
- Machine downtime vs revenue impact.

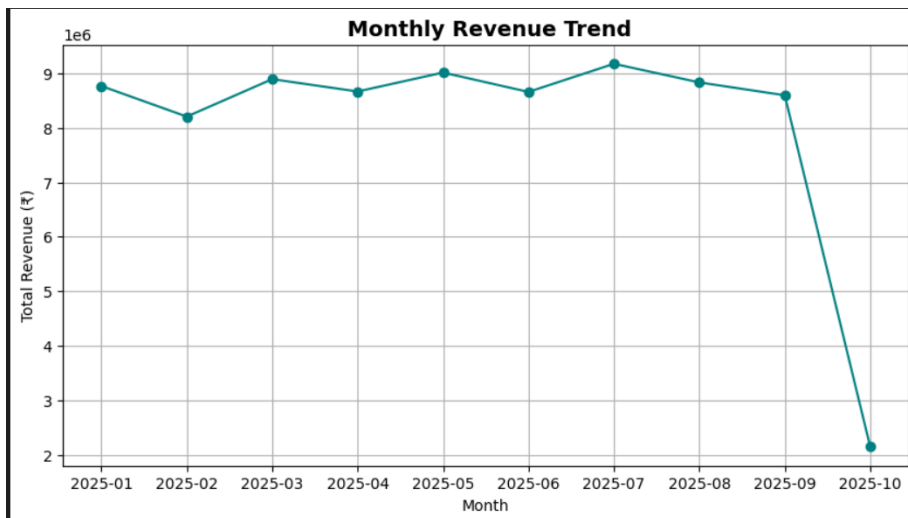
Monthly Revenue Trend:

```
Click to add a breakpoint efill_datetime"].dt.to_period("M") #
monthly_revenue = df.groupby("month")["revenue_generated"].sum().sort_index()

✓ 0.0s

plt.figure(figsize=(10,5))
plt.plot(monthly_revenue.index.astype(str), monthly_revenue.values, marker='o', color='teal')
plt.title(" Monthly Revenue Trend", fontsize=14, fontweight='bold')
plt.xlabel("Month")
plt.ylabel("Total Revenue (₹)")
plt.grid(True)
plt.show()

✓ 0.1s
```



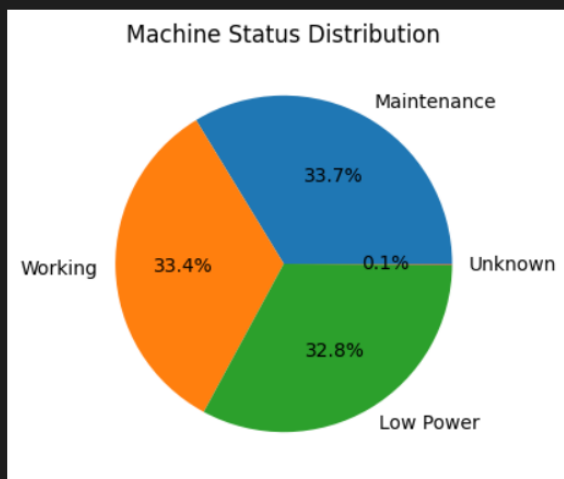
Machine downtime vs revenue impact.

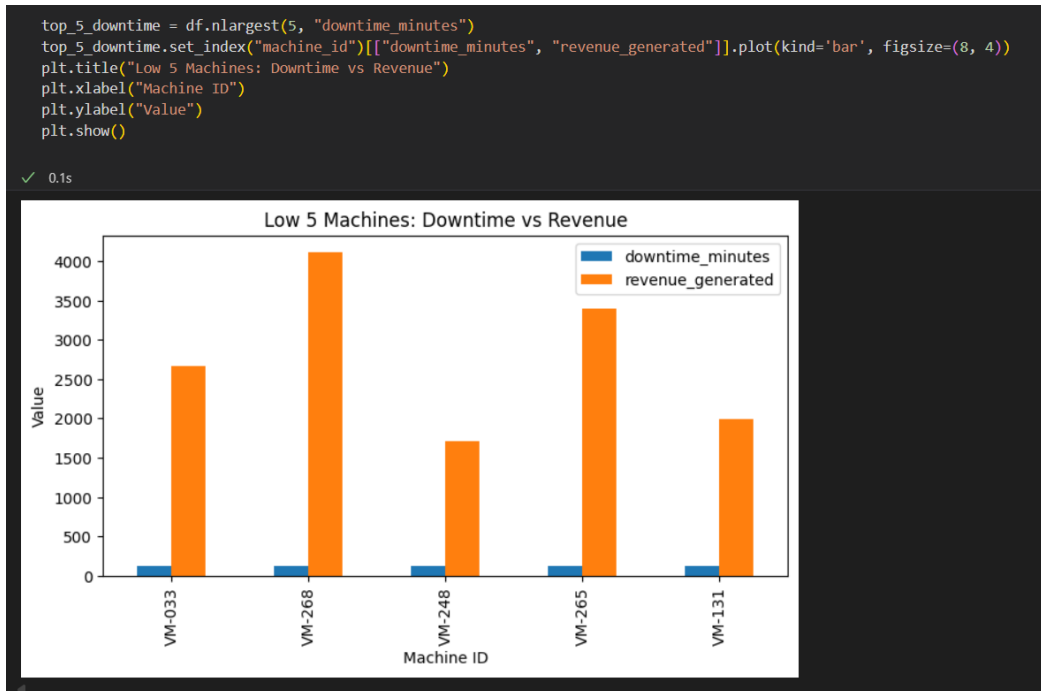
```
status = df["machine_status"].value_counts()
print(status)
plt.figure(figsize=(6,4))
status.plot(kind="pie", autopct="%1.1f%%", title= "Machine Status Distribution", ylabel='')
```

✓ 0.0s

```
machine_status
Maintenance    3372
Working        3340
Low Power      3281
Unknown         7
Name: count, dtype: int64
```

<Axes: title={'center': 'Machine Status Distribution'}>





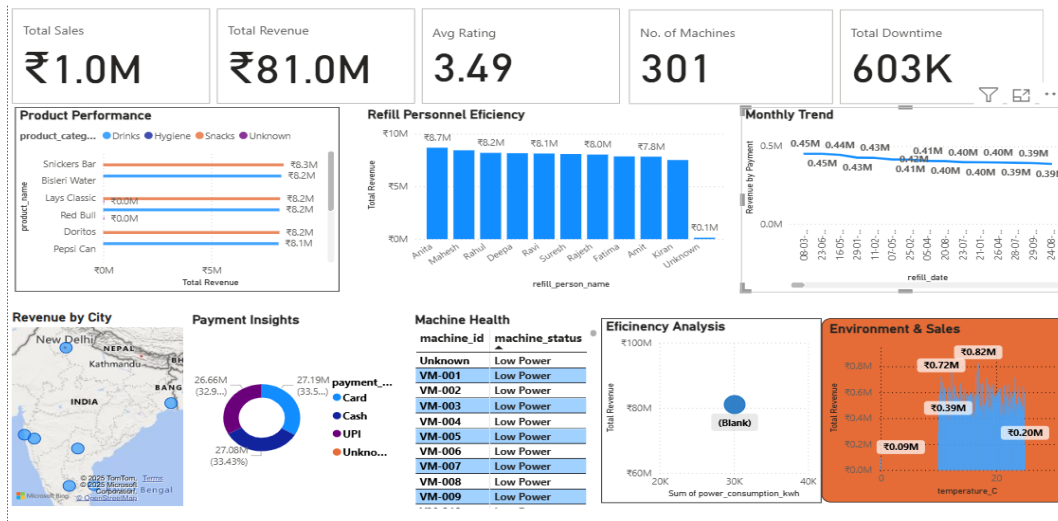
6. Dashboard Development

A dynamic Power BI dashboard was developed to visualize key vending machine KPIs. The dashboard connects to the cleaned data source and updates automatically on data refresh.

Key KPIs and Visuals:

- Total Revenue
- Average Customer Rating
- Machine Uptime Percentage
- Refill Frequency
- Top Performing Cities & Products

Interactive slicers were added for city, product category, and time period selection.



7. Results & Insights

- Machines in metropolitan cities generate approximately 40% higher revenue compared to smaller towns.
- Drinks category is the most profitable, contributing to nearly 55% of total sales.
- Machines with frequent maintenance logs show 25% lower uptime and 30% less revenue.
- Refill personnel efficiency directly correlates with machine uptime and consistent revenue generation.

8. Challenges & Solutions

Challenge 1: Missing Refill Person Data

Solution: Filled using median values grouped by machine_id and location.

Challenge 2: Mixed Date Formats

Solution: Standardized all refill dates using Pandas `to_datetime()` conversion.

Challenge 3: Power BI Data Refresh Delays

Solution: Optimized Power BI model relationships and reduced unnecessary columns.

9. Tools & Technologies Used

Database: MySQL

Data Cleaning: Python, Pandas, NumPy

Visualization: Power BI, Matplotlib

Reporting: Power BI Export

10. Conclusion & Recommendations

The Vending Machine Data Analysis project successfully demonstrated data management, analysis, and visualization capabilities. The developed Power BI dashboard provides valuable operational insights for decision-makers. Future improvements could include integrating real-time IoT data streams and automating Power BI refresh workflows.

- Integrate live IoT data for real-time vending machine monitoring.
- Enhance Power BI automation with scheduled refresh and alerts.
- Develop machine performance prediction using ML algorithms.