

Zudio Sales Dataset

Data Analysis + Retail Stock Predictive model

1. Project Overview

This project focuses on analyzing Zudio retail sales data to understand store performance, customer buying patterns, and category-wise demand across different locations. The analysis involved extensive data cleaning, transformation, SQL-based querying, and dashboard creation.

2. Dataset Summary

The dataset contained **7,899 sales records** and **28 columns**, including:

- Store details (city, state, area size, staff count, store type)
- Product information (category, clothing type, price, quantity)
- Customer details
- Sales & profit values
- Monthly/seasonal trends
- Operational data such as parking, security features, and opening date

3. Exploratory Data Analysis

df.isnull().sum()	
Store	0
Country	0
State	0
City	0
Category	0
Clothing Type	0
Store Number	0
Postal Code	0
Store Type	0
Store Open Date	0
Selling Area Size (sq ft)	0
Store Manager	0
Manager ID	0
Store Address	0
Contact Information	0
Operating Hours	0
Staff Count	0
Parking Availability	0
Security Features	2038
Order ID	0
Order Date	0
Month	0
Customer ID	0
Customer Name	0
Product ID	0
Price	0
Quantity	0
Sales Profit	0
	dtype: int64

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7899 entries, 0 to 7898
Data columns (total 28 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Store            7899 non-null    object  
 1   Country          7899 non-null    object  
 2   State             7899 non-null    object  
 3   City              7899 non-null    object  
 4   Category          7899 non-null    object  
 5   Clothing Type    7899 non-null    object  
 6   Store Number     7899 non-null    int64  
 7   Postal Code      7899 non-null    int64  
 8   Store Type        7899 non-null    object  
 9   Store Open Date   7899 non-null    object  
 10  Selling Area Size (sq ft) 7899 non-null    int64  
 11  Store Manager     7899 non-null    object  
 12  Manager ID       7899 non-null    object  
 13  Store Address     7899 non-null    object  
 14  Contact Information 7899 non-null    object  
 15  Operating Hours   7899 non-null    object  
 16  Staff Count       7899 non-null    int64  
 17  Parking Availability 7899 non-null    object  
 18  Security Features 5861 non-null    object  
 19  Order ID          7899 non-null    int64  
 20  Order Date         7899 non-null    object  
 21  Month             7899 non-null    object  
 22  Customer ID       7899 non-null    int64  
 23  Customer Name     7899 non-null    object  
 24  Product ID        7899 non-null    int64  
 25  Price              7899 non-null    int64  
 26  Quantity           7899 non-null    int64  
 27  Sales Profit      7899 non-null    float64 
dtypes: float64(1), int64(9), object(18)
memory usage: 1.7+ MB
```

```
df['order_date']=pd.to_datetime(df['order_date'])
```

```
df['order_date'].dtype
```

```
dtype('datetime64[ns]')
```

```
df.columns = df.columns.str.strip().str.lower().str.replace(' ', '_')

df.columns

Index(['store', 'country', 'state', 'city', 'category', 'clothing_type',
       'store_number', 'postal_code', 'store_type', 'store_open_date',
       'selling_area_size_(sq_ft)', 'store_manager', 'manager_id',
       'store_address', 'contact_information', 'operating_hours',
       'staff_count', 'parking_availability', 'security_features', 'order_id',
       'order_date', 'month', 'customer_id', 'customer_name', 'product_id',
       'price', 'quantity', 'sales_profit'],
      dtype='object')
```

- Certain cities showed consistently high revenue
- Women's clothing and T-shirts/Jeans sold the most
- Sales increased during festive months
- Store size and staff count showed mild correlation with quantity sold

4. Data Analysis using SQL

```
from sqlalchemy import create_engine

username='postgres'
password='root123'
host='localhost'
port='5432'
database='zudio'

engine = create_engine(f'postgresql+psycopg2://{{username}}:{{password}}@{{host}}:{{port}}/{{database}}')

table_name = 'sales'

df.to_sql(table_name, engine, if_exists='replace', index=False)

print(f'Data successfully loaded into table {table_name} in database {database}')


Data successfully loaded into table sales in database zudio
```

Object Explorer Dashboard Properties SQL Statistics Dependencies Dependents Processes Hey.sql zudio/postgres@PostgreSQL 18*

Servers (1)
 PostgreSQL 18
 Databases (3)
 postres
 Casts
 Catalogs
 Event Triggers
 Extensions
 Foreign Data Wrappers
 Languages
 Publications
 Schemas
 Subscriptions
 test
 zudio
 Casts
 Catalogs
 Event Triggers
 Extensions
 Foreign Data Wrappers
 Languages
 Publications
 Schemas
 Subscriptions

Login/Group Roles (17)
 pg_checkpoint
 pg_create_subscription
 pg_database_owner
 pg_execute_server_program
 pg_maintain

Query Query History Scratch Pad

```

1 select * from sales limit 20
2 --revenue generated by different category
3 select category, sum(sales_profit) as revenue
4 from sales
5 group by category
6 --average order value by category
7 select clothing_type, avg(sales_profit) as "Avg rev as cat"
8 from sales
9 group by clothing_type
10 order by avg(sales_profit) desc
11 --top 10 customers as per revenue
12 select customer_name,sales_profit
13 from sales
14 order by sales_profit desc
15 limit 5;
16 --statewise profit
17 select sales_profit
18 from sales
19 group by (state);
20
21
22
  
```

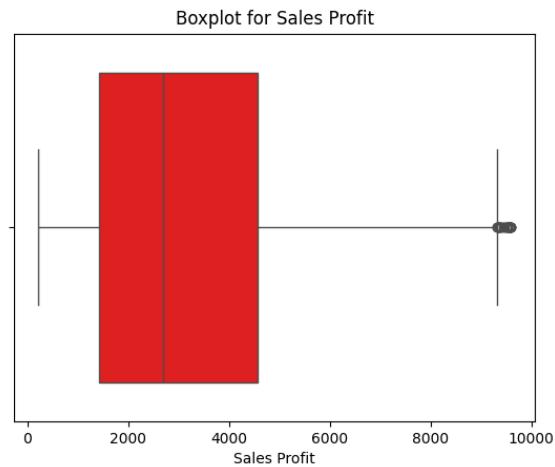
Data Output Messages Notifications SQL

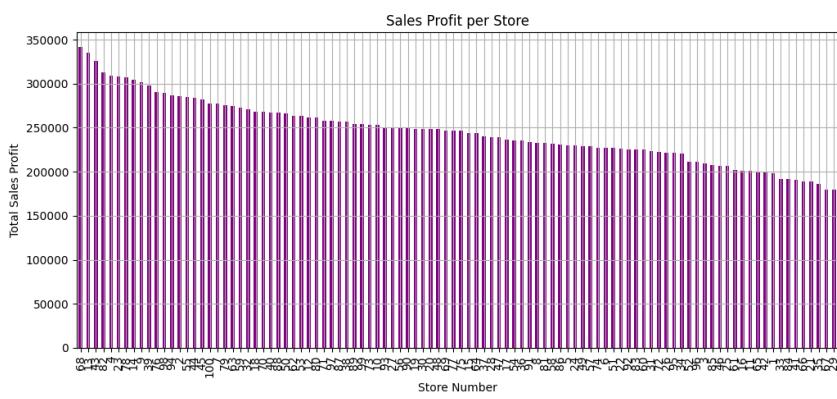
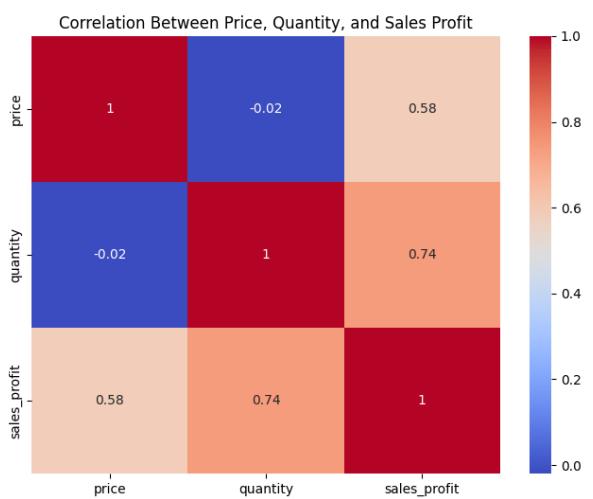
Showing rows: 1 to 20 Page No: 1 of 1 < << > >>

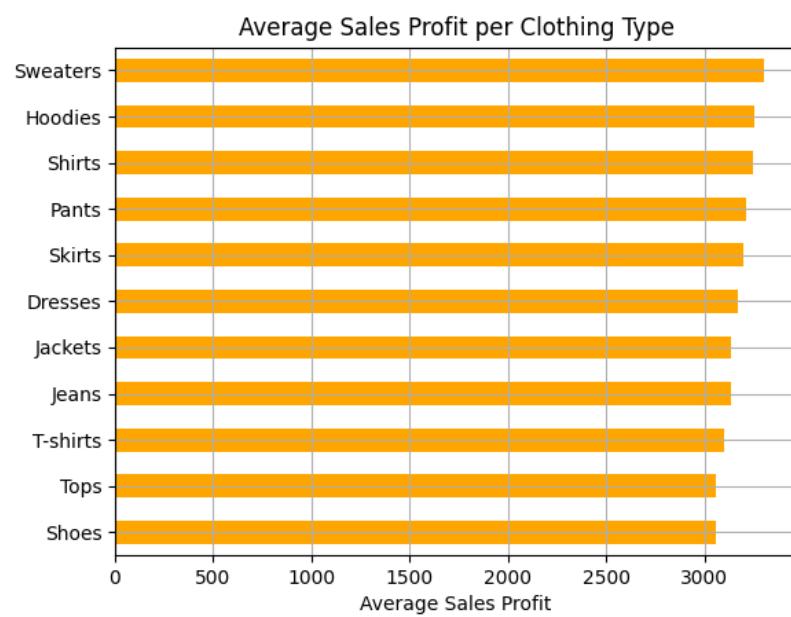
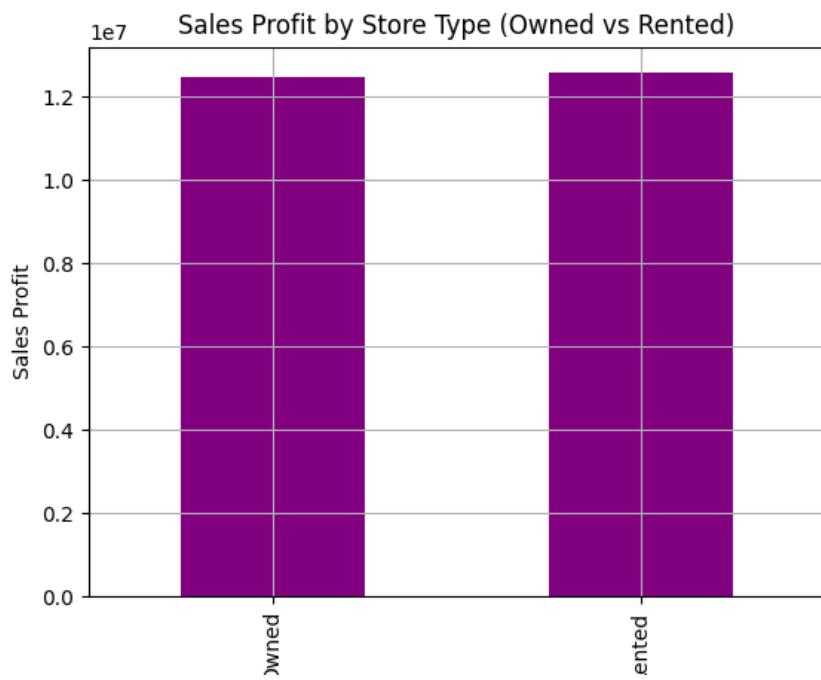
dng_availability	security_features	order_id	order_date	month	customer_id	customer_name	product_id	price	quantity	sales_profit
Available	Alarm	88240	2024-11-06 00:00:00	November	233084	Michele Ochoa	354	1958	5	3916
Available	Alarm	78145	2024-06-17 00:00:00	June	926472	Daniel Jimenez	741	2078	1	831.2
Available	Normal	88240	2024-11-06 00:00:00	November	233084	Michele Ochoa	354	1958	5	3916

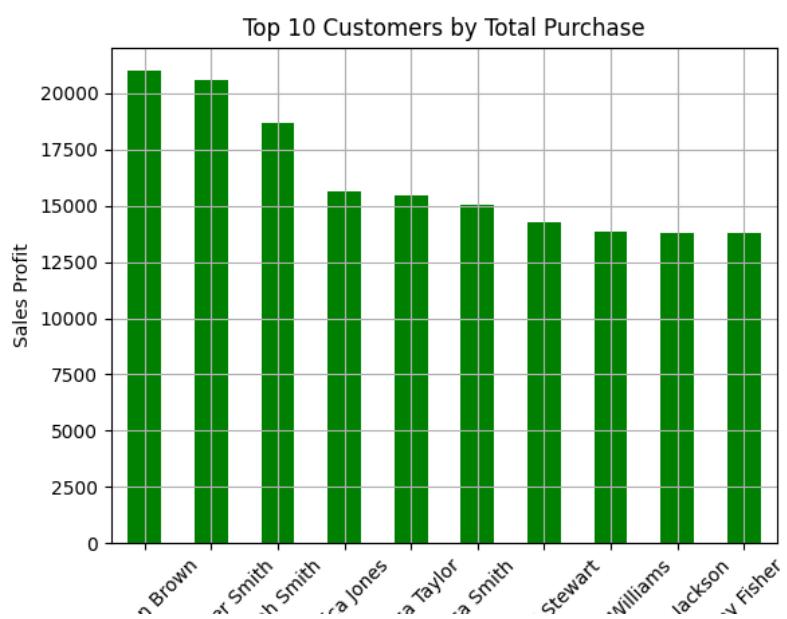
Total rows: 20 Query complete 00:00:00.225

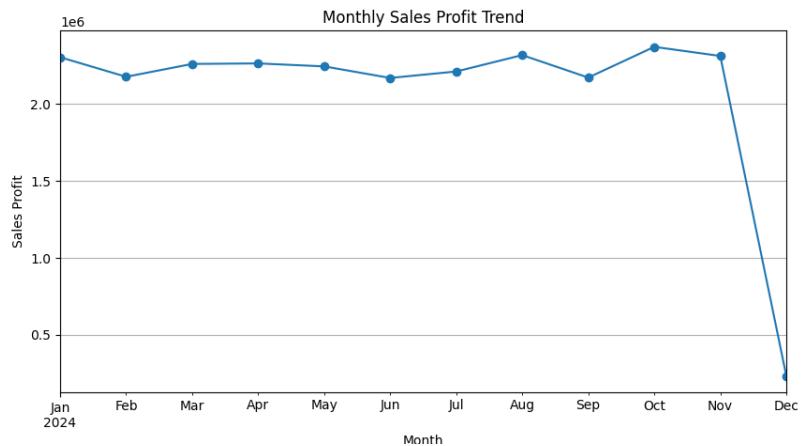
5. Data Visualisation



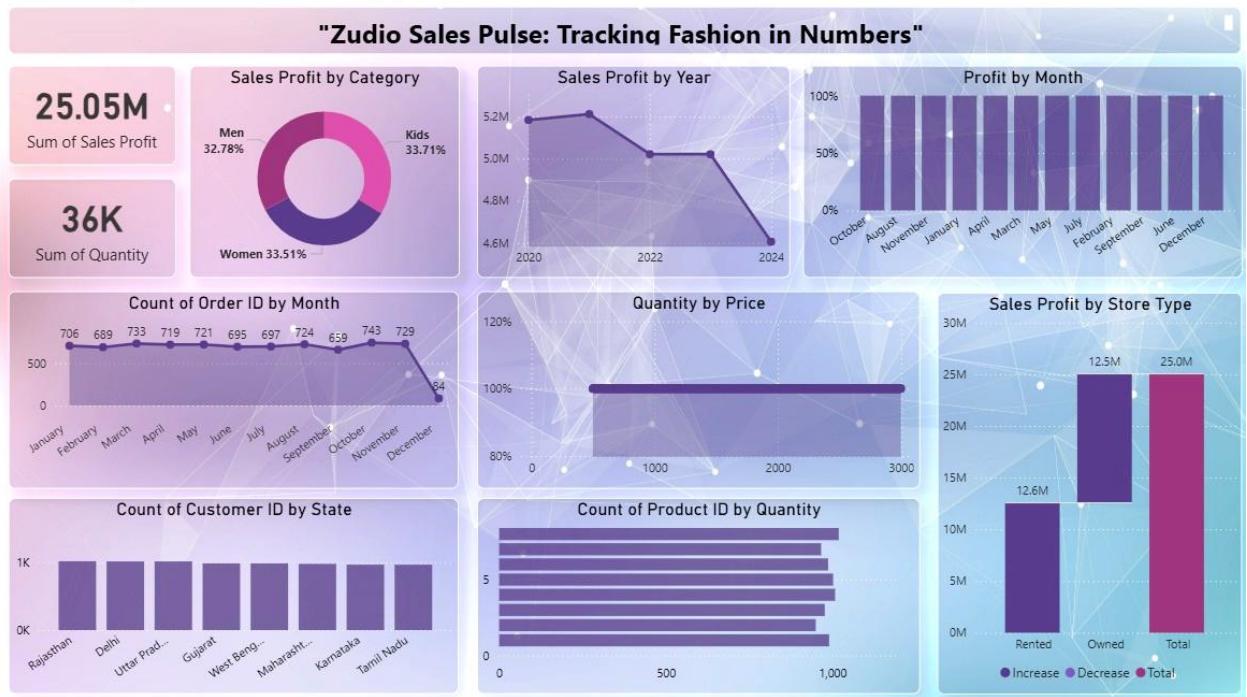








6. Power BI (Dashboard)



7. Predictive Model

-: Zudio Sales Quantity Predictor :-

Enter the following detail and system will predict the quantity that can be sold with respect to the input.

City

Category

Clothing Type

Store Type

Selling Area Size (sq ft)

Staff Count

Parking Availability

Month

Price Per Unit

Predicted Quantity Per Customer

Flag

Clear **Submit**

Use via API ↗ · Built with Gradio ➔ · Settings ⓘ

-: Zudio Sales Quantity Predictor :-

Enter the following detail and system will predict the quantity that can be sold with respect to the input.

City

Category

Clothing Type

Store Type

Selling Area Size (sq ft)

Staff Count

Parking Availability

Month

Price Per Unit

Predicted Quantity Per Customer

Flag

Clear **Submit**

Use via API ↗ · Built with Gradio ➔ · Settings ⓘ

Methods used:

- Feature preprocessing (Label Encoding, OneHotEncoding, imputation)
- Train-test split
- Model experimentation with:
 - Random Forest Regressor
 - Gradient Boosting (HistGradientBoostingRegressor)
 - Linear Regression

