**1. Problem Definition**

**Key Questions:**

* What are the sales trends and which products/stores are driving revenue?
* How do customer demographics and behavior vary by store location?
* What external factors (e.g., weather, holidays, demographics) are impacting sales?
* How can RetailPulse improve sales and optimize product placement?

**Primary KPIs to Track:**

1. **Sales Metrics**:
   * Total Revenue
   * Sales Growth Rate
   * Sales by Category/Location
2. **Customer Metrics**:
   * Average Spend per Customer
   * High Spenders vs. Low Spenders
3. **Operational Metrics**:
   * Store Performance (e.g., Revenue by Store, Geographic Trends)
   * Category Performance (e.g., Top-performing vs. Low-performing products)
4. **External Factors**:
   * Holiday Sales Lift
   * Weather Impact (e.g., rainy days vs. sunny days)

**2. Objective**

The goal of this project was to analyze sales and customer behavior across multiple retail stores, gain insights into sales trends, understand the drivers of high and low sales, and assess the impact of external factors like holidays and weather on sales. Using this analysis, actionable recommendations were made to improve overall sales performance and optimize product placement.

**3. Datasets**

**1.Data Set Creation :**

Downloaded partial sales dataset from “Kaggle” : <https://www.kaggle.com/datasets/mohammadtalib786/retail-sales-dataset>.

**2.Adding Store data:** The dataset didn’t had any sales store information, so had to enrich it using real-world knowledge and randomly assigning them to each transactions on the dataset. Generated the enriched\_data.

“store.csv”

Store\_ID,Location,Size,Category

S001,New York,Large,Flagship

S002,Los Angeles,Medium,Outlet

S003,Chicago,Small,Online

S004,Texas,Large,Outlet

S005,Atlanta,Medium,Flagship

S006,Miami,Large,Online

S007,Colorado,Small,Outlet

S008,San Diego,Large,Flagship

S009,Oregon,Small,Outlet

S010,New Orleans,Large,Online

**3. Feature Engineering:**

* Added new features on date column to reveal insights and identify trends.
* Calculated ‘Sales per Customer’ and ‘Repeat Purchase’ columns for future use.

**4.Integrating Weather Data into the dataset**

* Used OpenWeatherMap API to get data for each transactions based on ‘Location’ and ‘Date’.
* Calculated coordinates (latitude and longitude) using geopy and then merged them into the dataset. Defined a ‘fetch\_weather’ function which used these coordinates and date to extract data from json response after we call the API. Added a loop to input the weather data into the sales dataset.

**5.Adding Holidays using Caledarific API:**

* Using calendarific API, defined a ‘fetch\_holidays’ function to get response based on country (which is US) and year (extracted from each date on sales table) to fetch if that particular date is a holiday or not and the ‘holiday\_name’. Added two new columns named ‘is\_holiday’ and ‘holiday\_name’ into the dataset.

**6.Studied Correlation of Weather and Holidays on Sales:**

* Using a scatterplot to study the impact of holidays and weather on sales trend.

**7.Holiday Sales Lift:**

* Compared sales on holidays vs normal days to understand the effect on sales trend.

**8.Data Cleaning Steps:**

* Checked for any NULL values, duplicates, invalid dates.
* Standardized the date column since we would be using that for adding features and adding Holidays and getting Weather information.
* Normalized the ‘location’ and ‘weather’ columns.
* Created a boxplot to identify outliers in the ‘Total Amount’.

**4. Key Findings**

**1. Sales Trends**

* **Seasonality Impact:** Sales showed strong seasonality, with revenue peaks during holidays like Christmas, New Year, and Thanksgiving.
* **Time-Based Performance:** Weekends & Holidays consistently had higher sales compared to weekdays, likely due to increased foot traffic.

**2. Customer Segmentation**

* **High-Value Customers:** A small segment of customers (e.g., top 20%) contributed to a significant portion (e.g., 60-80%) of total revenue (Pareto principle).
* **Repeat Customers:** Repeat customers had higher average spend compared to one-time buyers, highlighting the importance of customer retention.

**3. External Factors**

* **Holidays:** Revenue saw a significant boost during holidays, with holiday promotions driving an average sales increase of **17%** compared to non-holiday periods. This emphasizes the importance of holiday-specific promotions.
* **Weather Impact:**
  + Sunny days correlated with higher in-store sales due to increased foot traffic.
  + Rainy or snowy days saw a drop in foot traffic but an increase in online sales for certain categories like groceries and electronics.

**4. Geographic and Store Performance**

* **High-Performing Locations:** Easy to identify high and low performing locations based on the data which can help in creating more location specific marketing strategies.
* Top-performing stores can act as benchmarks. Their strategies and practices can be analyzed and replicated in other locations to improve overall performance.

**5. Forecasting & Planning**

* Recognizing monthly sales patterns lays the groundwork for more accurate sales forecasting, ensuring that resources are allocated efficiently.
* With this forecast, we can adjust inventory levels and marketing strategies to match expected demand. For example, we can ramp up promotions during months with anticipated sales spikes.

1. **Challenges**

**1. Data Quality Issues**

* **Inconsistent Formats:** Data formatting issues (e.g., different date formats, inconsistent product categories) slowed down preprocessing.

**2. External Data Integration**

* **Weather Data:** Integrating weather data for multiple locations and matching it with corresponding dates posed a challenge. Historical weather data retrieval was limited to a specific time range.
* **Holiday Data:** Identifying regional and global holidays required external data sources, some of which lacked complete coverage or required manual adjustments.

**3. Statistical Challenges**

* **Noise in Data:** Certain stores had high variability in sales performance due to temporary promotions or outlier events, making it difficult to generalize findings.
* **Correlation vs. Causation:** While correlations were identified (e.g., between weather and sales), proving causation required more advanced analysis.

**4. Tool and Performance Limitations**

* **API Rate Limits:** External APIs for weather and holiday data had strict rate limits for free tier as well as historical data limitations.
* **Scalability:** Processing large datasets (e.g., 1,000+ rows) while performing advanced analyses (e.g., regression, clustering) required optimizing code and leveraging efficient libraries like NumPy or Pandas.

**6. Suggestions for Improving Sales**

**1.Targeted Marketing:**

* + Offer discounts to high-value customer segments.
  + Run location-specific campaigns based on store demographics.

**2.Seasonal Promotions:**

* + Create targeted promotions around holidays.
  + Adjust inventory based on weather forecasts (e.g., umbrellas during rainy days).

**3.Low-Performing Store Strategy**:

* + Analyze staffing, inventory, and layout issues.
  + Increase advertising in low-performing store regions.

**7. Conclusion**

This project successfully identified actionable insights into sales trends, customer segmentation, and external factors affecting revenue. Retailers can apply these findings to improve marketing strategies, optimize inventory, and drive long-term growth.

**Tableau Dashboard:** Link:https://public.tableau.com/app/profile/sushant.k8220/viz/RetailSalesDashboard\_17373267818180/RetailPulse