

DOCUMENTATION

1. Business Problem and Motivation

E-commerce sales in the United States experience fluctuations throughout the year. Retailers often rely on holiday seasons to boost revenue, but it's not always clear which holidays drive the most sales or where demand is highest. The goal of this project is to explore how U.S. holidays impact sales across time, region, and product category.

This analysis seeks to answer:

1. Do holidays lead to increased sales or profit?
2. Which states and cities are most active during holidays?
3. Which product types perform better on holidays?

Understanding these trends helps businesses forecast demand, optimize inventory, and schedule promotions more effectively during holiday periods.

2. Description of Datasets and Cleaning Process

Two datasets:

- Main dataset: United States E-Commerce Records 2020 (from Kaggle)
Includes ~10,000 transaction records with details such as order date, product name, category, price, quantity, state, city, and profit.

Link: <https://www.kaggle.com/datasets/ammaraahmad/us-ecommerce-record-2020>

- Supplemental dataset: U.S. Holidays 2020 (scraped from Timeanddate.com)
Contains holiday names, types, and dates across 2020.

Link: <https://www.timeanddate.com/holidays/us/2020>

Data Preparation & Cleaning in Power Query included:

1. I web scrapped data about holiday date's and modify,clean and make more readable by python
2. Removing duplicates in the Holidays table (to ensure one row per date)
3. Merging the datasets using Order Date and Holiday Date
4. Modify formats of Price, date, discount, profit columns, checked correctness of other columns categories, and location names
5. Creating 3 calculated columns and 2 hierarchy's:
 - Total Price = Price × Quantity

- Day of Week
- isHoliday (Yes/No)

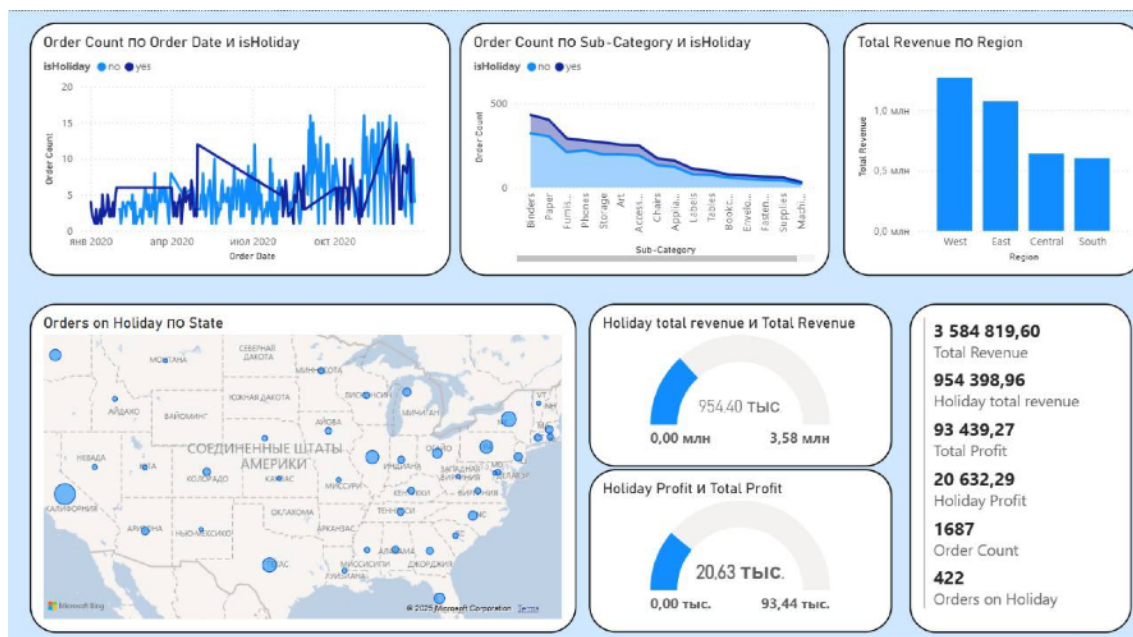
And

- Location: State, City and Postal Code
 - Category hierarchy: Category, Sub-Category, Product Name, and Product ID
6. Handling null or inconsistent values by Power Query
 7. Building relationships in a star schema model between «sales_merged» and «Holidays» by «Order Date» in Sales_merged and «Date» in Holidays

3. Summary of Insights

So my project shows that in our dataset:

- Out of 1,687 total orders, 422 were made on holidays (≈25%)
- Holiday sales totaled \$954,398.96 (≈27% of total revenue)
- Holiday profit was \$20,632.29 out of \$93,439.27 (≈22% share)
- Top-performing states during holidays: California, New York, Texas
- Holiday-related spikes in order volume were observed in late November (Thanksgiving) and December (Christmas)
- Categories like Phones and Binders had notably higher demand during holidays



Also I have, 6 visualizations that help to understand our dataset and see the difference between regular day and Holidays.

1. Visualizations shows us «Total Values» of all purchases and revenue, profits that will help us. It is «Total Revenue», «Holiday Total Revenue», «Total Profit», «Holiday profit», «Order Count», «Orders on Holiday»
2. It is visualization that compares values of regular days and Holidays, like Total revenue vs Holiday Total revenue, All orders vs Holiday orders, Total Profit and Holiday Profits.
3. Is a Linear Graphs, that shows us how Holidays affect on Orders by Month

4. It is a graphics, that shows us a TOP products that sells in Holidays and in regular days, and graph also shows their differences.
5. It's a Map of USA which shows us in which states are high and low purchases in US by size of circles(Big circles = a lot of purchases, Small circles = a low purchases).
6. It is a Histogram which shows us in which regions of US had the most and the less revenue.

4. **Limitations of the Analysis**

- Holiday purchases may occur before the actual holiday date, but the current model only captures activity on the exact holiday. As a result, the analysis might underestimate the true impact of holidays on sales.
- Treating all holidays equally (without distinguishing between major, minor, regional, or cultural observances) can dilute the importance of key holidays like Christmas or Thanksgiving and overemphasize less impactful dates.
- Regions (states or cities) with very few transactions may appear to have low or inconsistent sales activity. However, this may be due to data sparsity rather than actual customer behavior, which can mislead geographic comparisons.
- Without considering operating costs (such as logistics, warehousing, or procurement), profit figures may be overstated. High sales in holidays might not translate into high net profit if costs also rise during those periods.
- Excluding returns and cancellations means the analysis reflects gross profit only, not net. Some holidays might have higher return rates (e.g., post-Christmas), which would reduce actual profitability.
- Lack of data on sales channel or device limits the ability to optimize marketing or user experience. For example, if mobile purchases dominate during holidays, this insight would be missed — leading to underinvestment in mobile UX or app marketing.

5. **Recommendations Based on Findings**

Increase marketing efforts ahead of major holidays

Sales and profits significantly increase during holidays, especially in November and December. It is recommended to launch marketing campaigns 5–7 days before key holidays such as Thanksgiving, Black Friday, and Christmas.

Prepare logistics and inventory in high-demand regions

The states with the highest holiday sales volume are California, New York, and Texas. It is advisable to increase inventory levels and logistics readiness in these regions in advance of major holidays.

Create seasonal product offerings and bundles

Certain categories (e.g., electronics, office supplies, accessories) show increased demand during holidays. Businesses should develop special holiday bundles, discounts, or limited-edition products to boost conversion.

Prioritize holidays by commercial importance

Not all holidays drive the same level of demand. It is recommended to classify holidays by impact (e.g., federal, religious, commercial) and focus resources on those that generate the highest returns.