E- Commerce Recommendation System

Jay Shukla

College of Engineering
Northeastern University
Toronto, ON
Shukla.j@northeastern.edu

1 Definition of Problem

Overview: Typically, e-commerce datasets are proprietary and consequently hard to find among publicly available data. However, The UCI Machine Learning Repository has made this dataset containing actual transactions from 2010 and 2011. The dataset is maintained on their site, where it can be found by the title "Online Retail."

This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers.

As per the UCI Machine Learning Repository, this data was made available by Dr Daqing Chen, Director: Public Analytics group. chend '@' lsbu.ac.uk, School of Engineering, London South Bank University, London SE1 0AA, UK.

Kaggle Dataset: https://www.kaggle.com/datasets/carrie1/ecommerce-data?resource=download

2 Importing Data

First, we will start with importing the data.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
81943	543182	22435	SET OF 9 HEART SHAPED BALLOONS	4	2011-02-04 10:40:00	2.46	NaN	United Kingdom
128096	547249	85186A	EASTER BUNNY GARLAND OF FLOWERS	2	2011-03-22 09:27:00	0.83	NaN	United Kingdom
354688	567890	23375	50'S CHRISTMAS PAPER GIFT BAG	10	2011-09-22 15:40:00	0.82	14125.0	United Kingdom
223162	556474	22386	JUMBO BAG PINK POLKADOT	1	2011-06-12 12:01:00	2.08	16007.0	United Kingdom
251465	559055	21913	VINTAGE SEASIDE JIGSAW PUZZLES	1	2011-07-05 17:09:00	3.29	NaN	United Kingdom
275429	560991	21638	ASSORTED TUTTI FRUTTI NOTEBOOK	12	2011-07-22 13:29:00	2.10	12438.0	Norway
26031	538453	20666	ECONOMY HOLIDAY PURSE	1	2010-12-12 12:26:00	2.95	16779.0	United Kingdom
166048	550838	21931	JUMBO STORAGE BAG SUKI	2	2011-04-21 11:31:00	2.08	14577.0	United Kingdom
308681	564049	22536	MAGIC DRAWING SLATE PURDEY	10	2011-08-22 13:30:00	0.42	17585.0	United Kingdom
201140	554271	21380	WOODEN HAPPY BIRTHDAY GARLAND	6	2011-05-23 13:06:00	2.95	16422.0	United Kingdom

3 Understanding the Data

The dataset contains transactions for an e-commerce service, we can deduct the following about the attributes from our first glance:

- Invoice Number (InvoiceNo): A unique code for each transaction. If it starts with 'c', i guess it means the transaction was cancelled.
- Product Code (Stock Code): A unique code for each product item.
- Product Description (Description): The name of the product.
- Quantity: The number of each product sold in a transaction.
- Invoice Date (Invoice Date): The date and time when the transaction occurred.
- Unit Price (Unit Price): The price of one unit of the product in currency.
- Customer ID (CustomerID): A unique code for each customer.
- Country: The country where the customer resides.

4 Data Cleaning

First, we will display rows with missing values in the ‘description’ column.

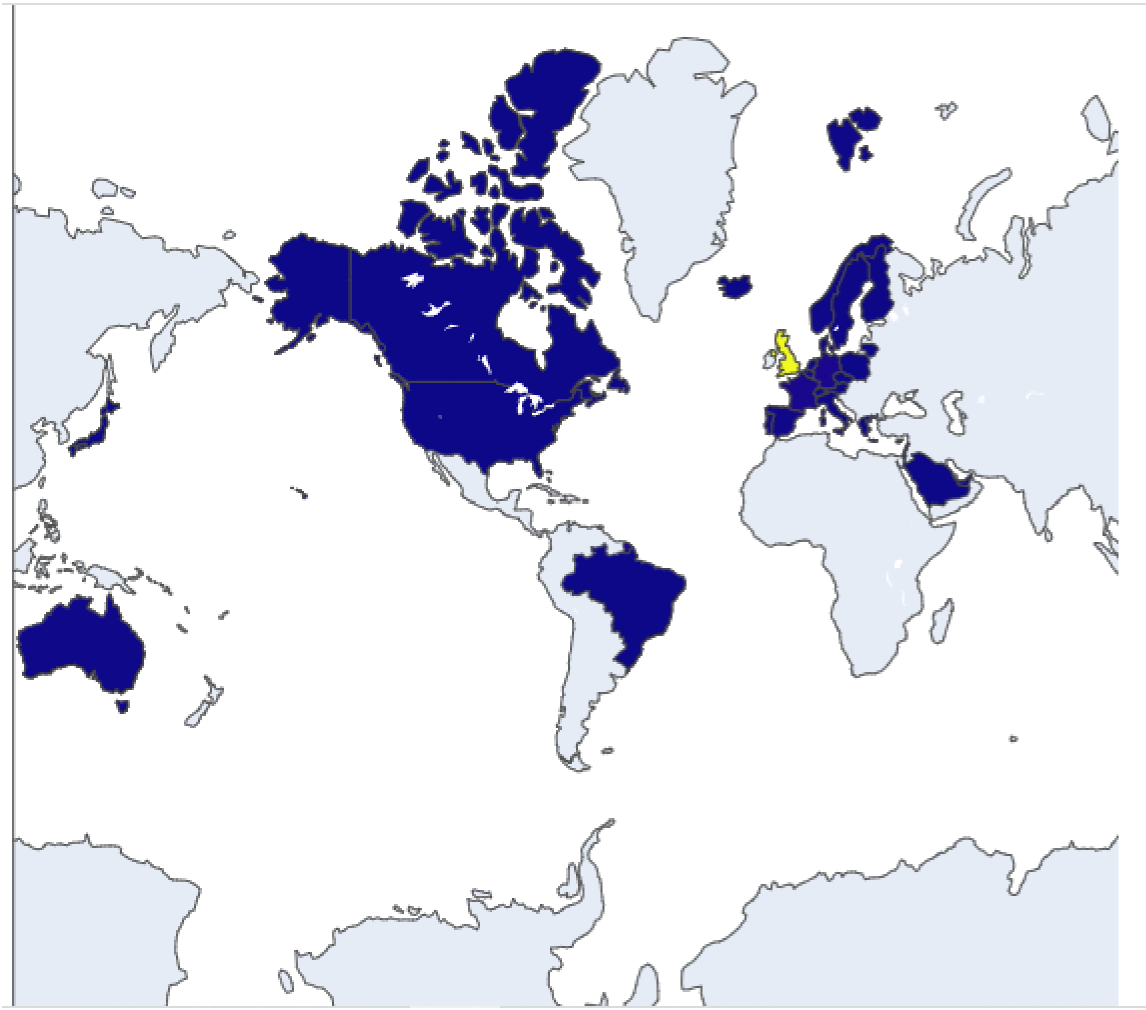
- I found 10062 data duplication.
- After that I can still see other two issues
- Unit price has 0 as the minimum value.
- Quantity has negative values.
- Number of occurrences where Unit price is 0.0: 40.
- After that we will correct the datatypes.
- As some Stock codes refer to different descriptions of the same items, let’s make sure to only have one description per stock code.

5 Data Exploration

We will start with time period of the data.

- The data contains a transactional history of little over a year.
- There is total 37 number of countries.

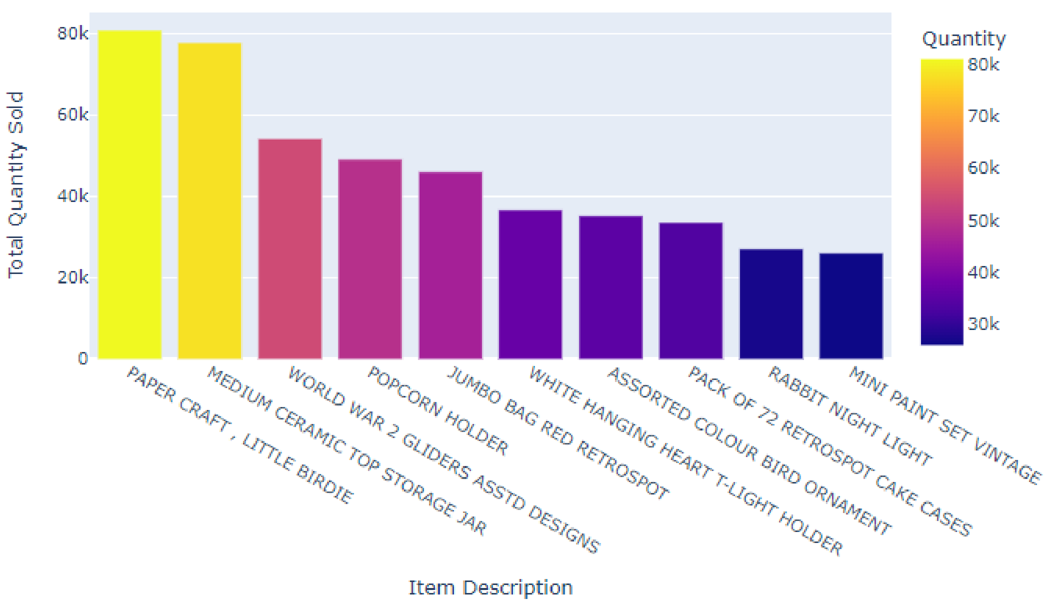
Total number of orders per country



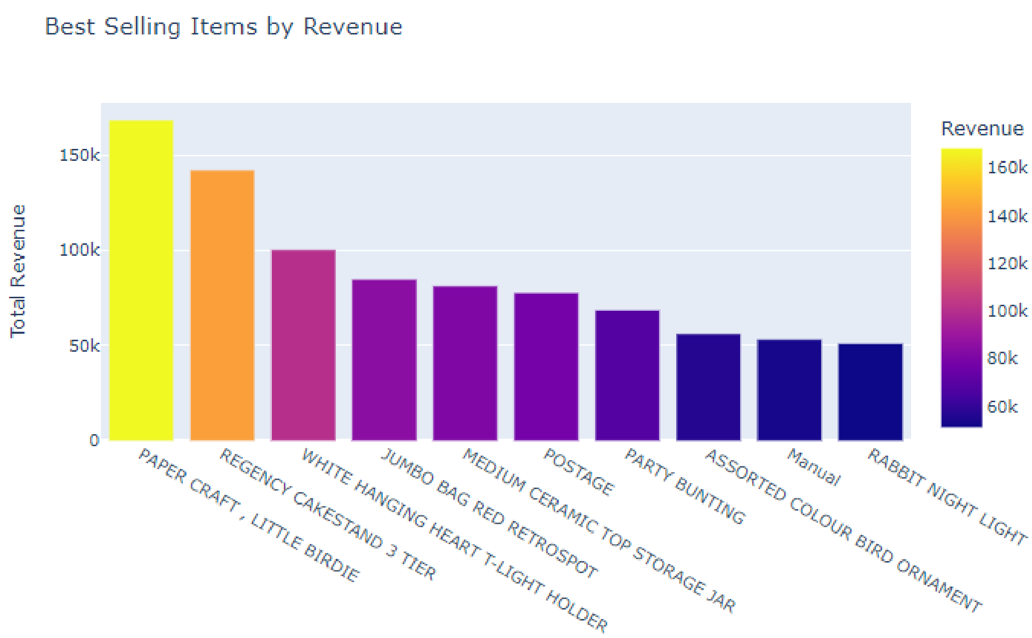
Highest Revenue Country

Here, we will calculate total revenue.

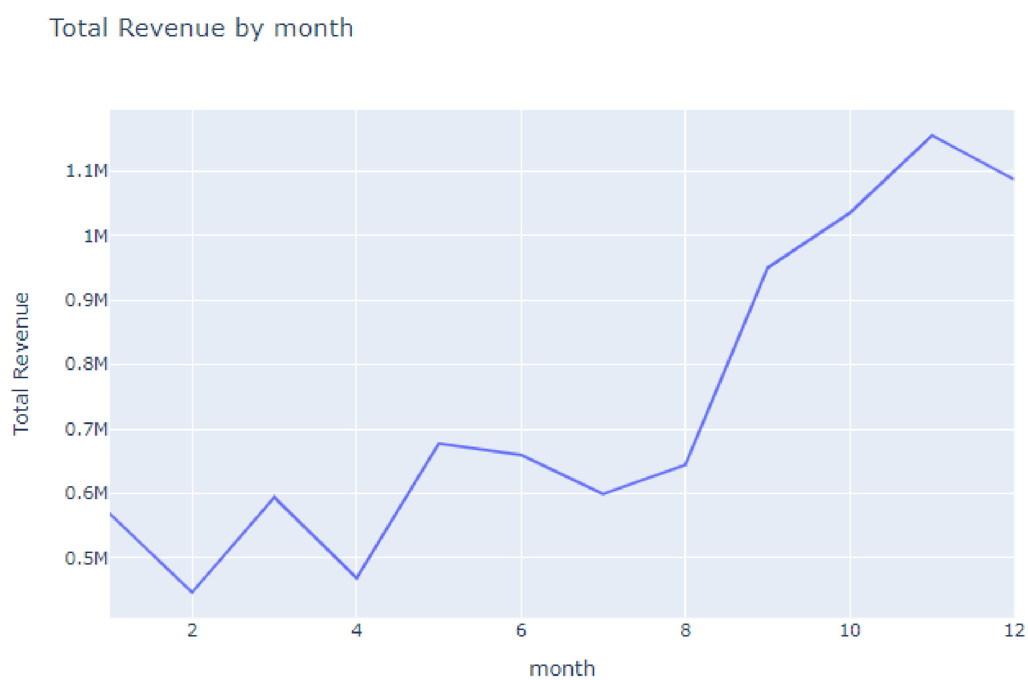
Best Selling Items by Quantity



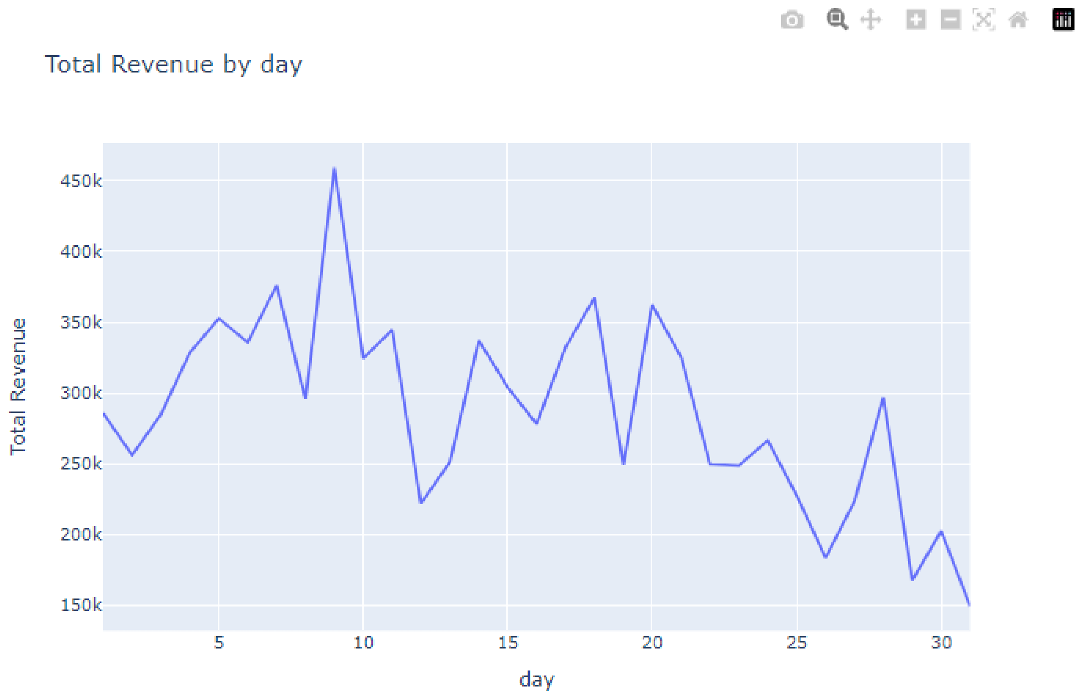
Best Selling Items by Revenue



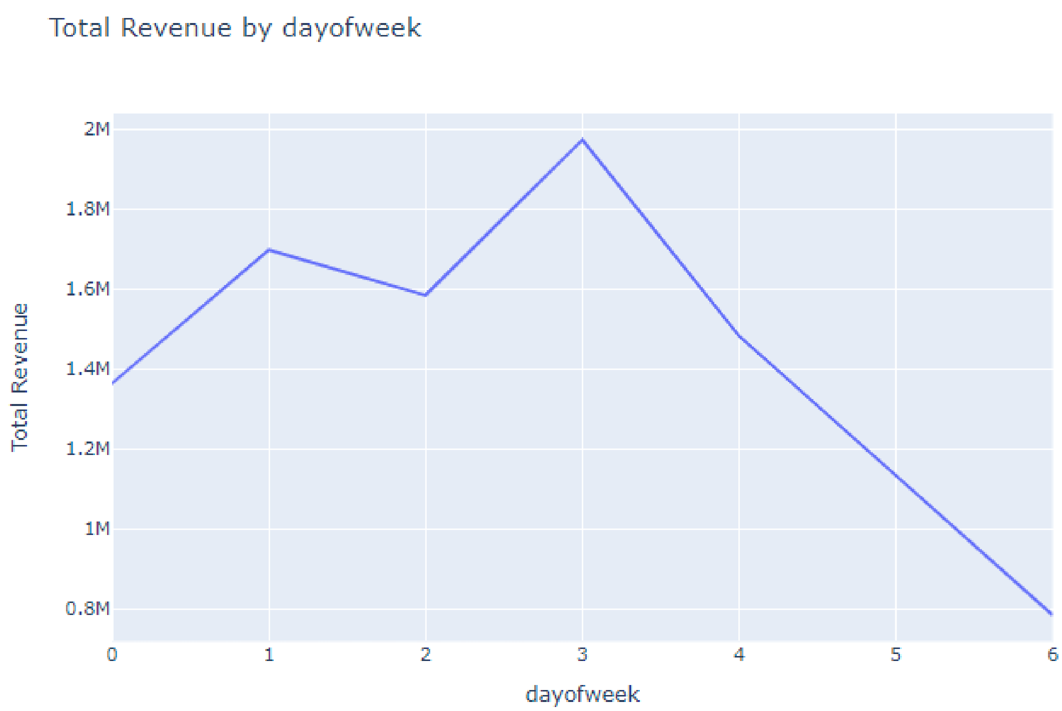
Total Revenue by Month



Total Revenue by Day

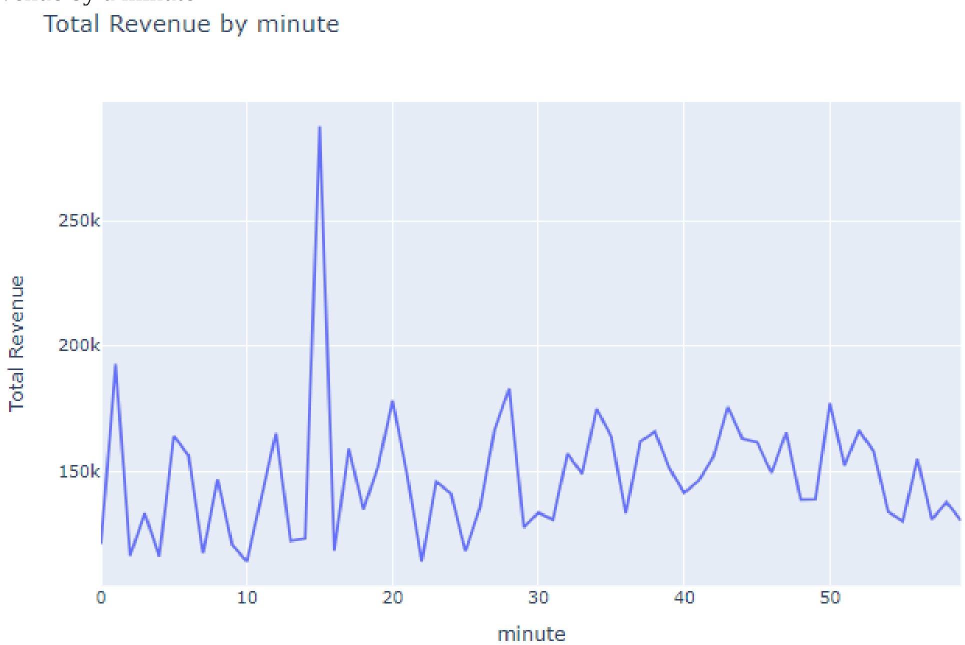


Total Revenue by day of week



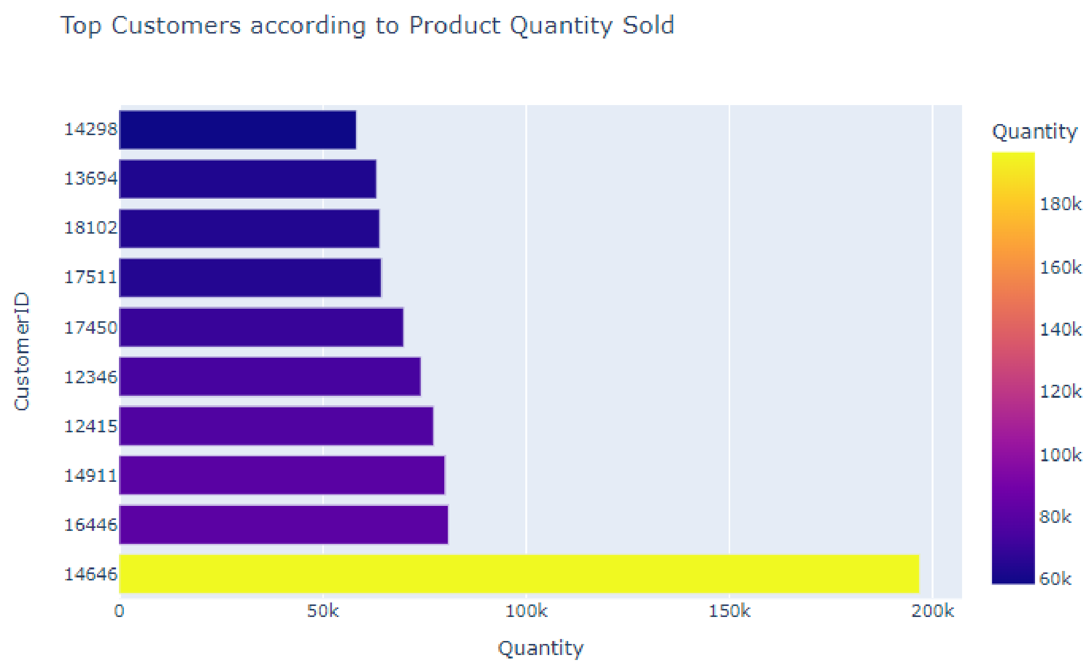
Total Revenue by an hour

Total Revenue by a minute

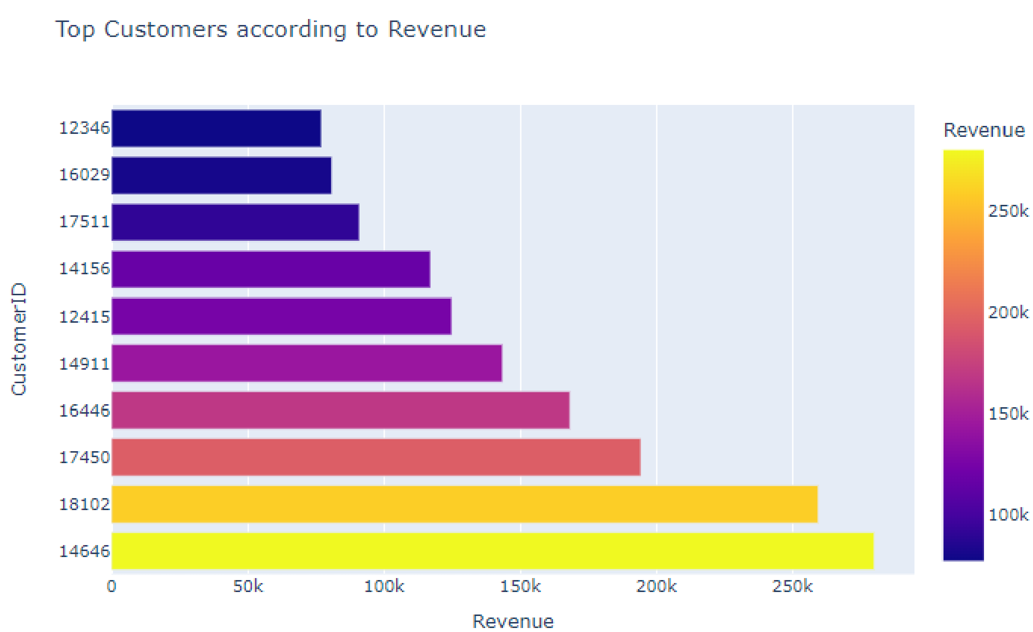


Now we will work on Number of Clients

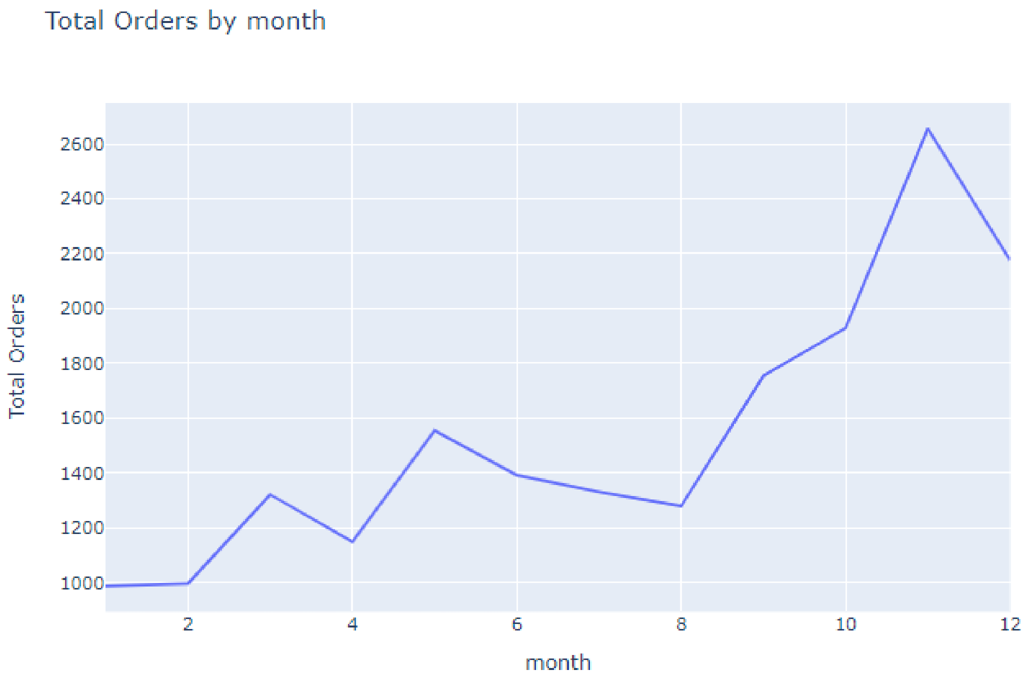
Top Customers according to product quantity sold.



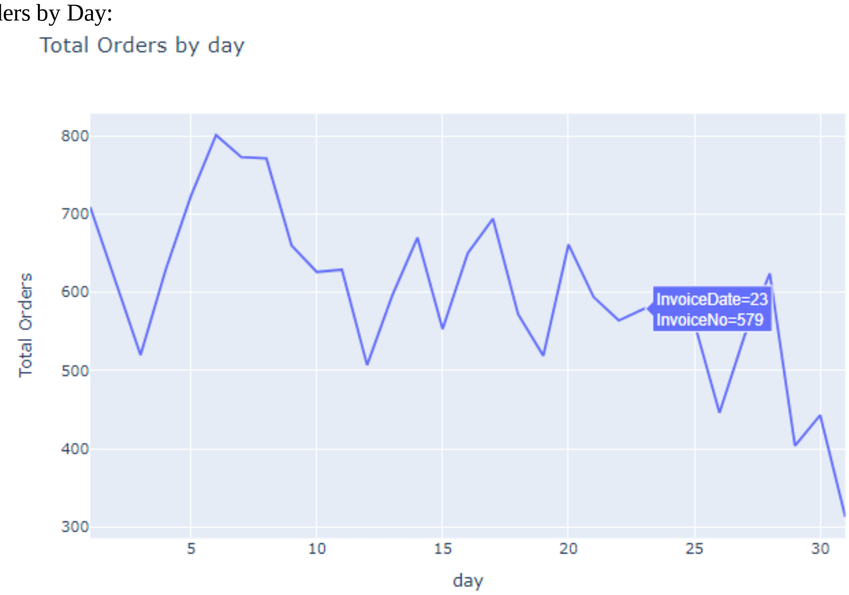
Top Customers According to Revenue



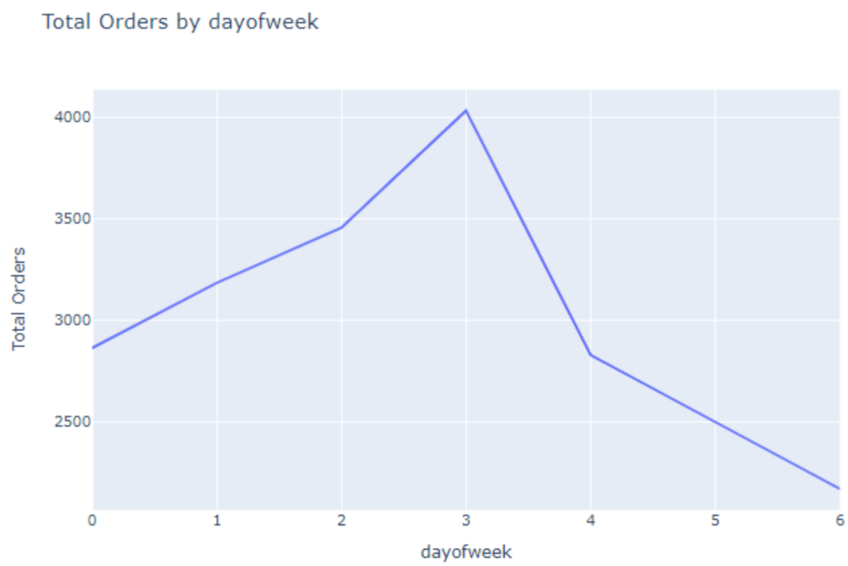
Total Order By Month:



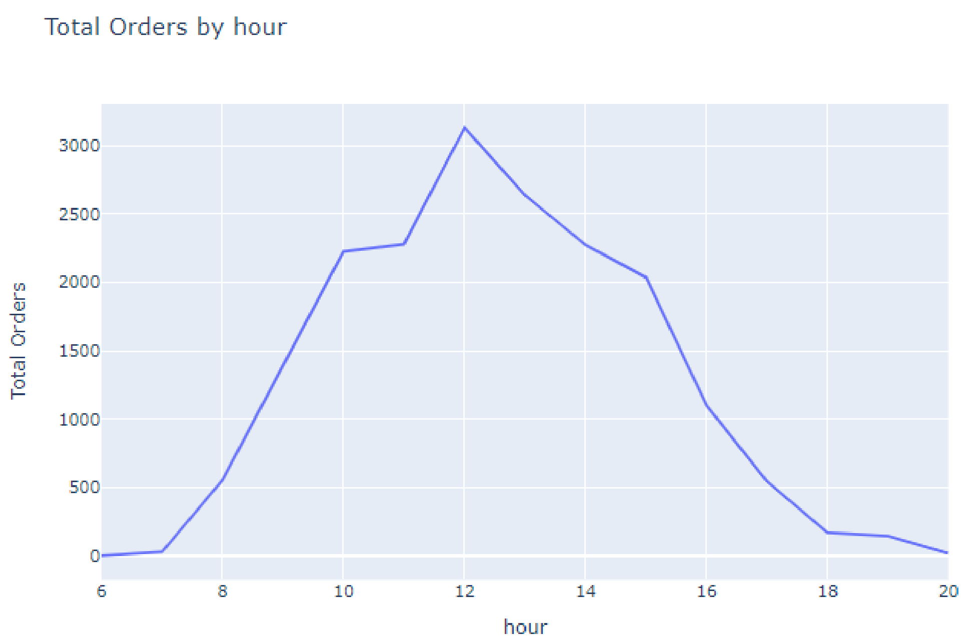
Total Orders by Day:



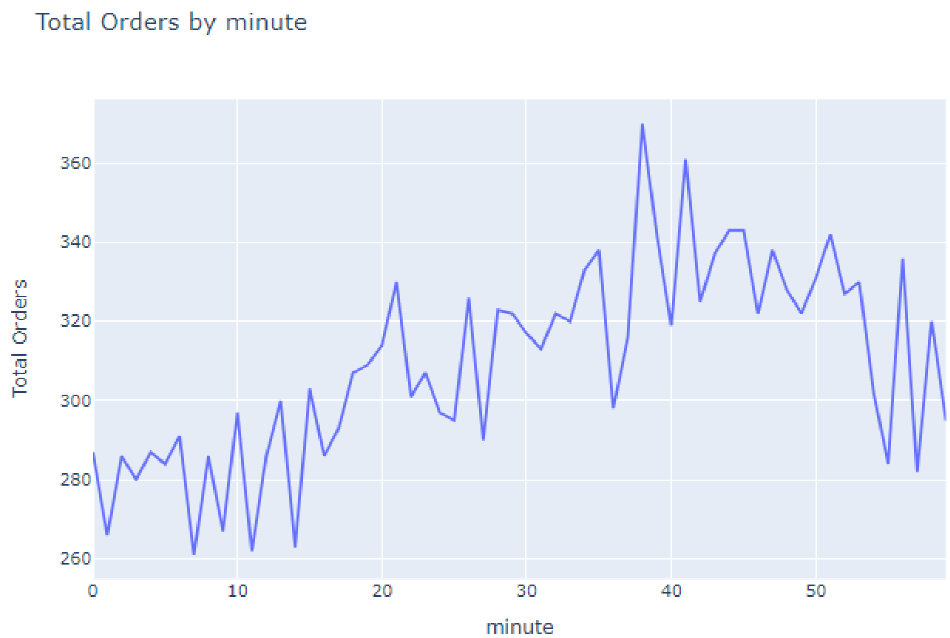
Total Order by Day of Week



Total Order by Hours



Total Order by Minutes



6. Recommendation for you

- So I will start with data transformation and after that we will apply cosine similarity.
- Here is the evaluation.
  Evaluation for Randomly Selected User ID: 13835

  Past Recommendations:
  1. PACK OF 72 RETROSPOT CAKE CASES
  2. TRAVEL CARD WALLET KEEP CALM
  3. HOMEMADE JAM SCENTED CANDLES
  4. WORLD WAR 2 GLIDERS ASSTD DESIGNS
  5. JUMBO BAG RED RETROSPOT

  Actual Future Purchases:
  1. MEMO BOARD COTTAGE DESIGN
  2. WHITE HANGING HEART T-LIGHT HOLDER
  3. NATURAL HANGING QUILTED HEARTS
  4. CHARLOTTE BAG SUKI DESIGN
  5. IVORY WICKER HEART LARGE
  6. 15CM CHRISTMAS GLASS BALL 20 LIGHTS

  Correct Recommendations:

  Metrics:
  Precision: 0.00
  Recall: 0.00

Overall Model Metrics:

Mean Average Precision: 0.34
Total Precision: 0.16
Total Recall: 0.02

Based on the evaluation results, we can gain insights into the strengths and weaknesses of the recommendatio n engine:

Strengths:

Mean Average Precision (MAP): The Mean Average Precision measures the overall quality of the recommen dations across all users. In our case, the MAP is 0.34, which indicates that, on average, the recommendations are somewhat accurate and relevant.

Precision: Precision is the proportion of correct recommendations out of all recommendations made. The tota l precision of 0.16 suggests that when recommendations are made, there's a reasonable chance that some of th em are correct.

Weaknesses:

Recall: Recall is the proportion of correct recommendations out of all the items that a user actually purchased in the future. The total recall of 0.02 indicates that the engine has a low ability to capture all the items a user might purchase in the future.

Mean Average Precision (MAP): While the MAP is relatively decent, it's not extremely high. This suggests t hat the engine's ability to recommend items that the user will purchase in the future is still moderate.

Overall Performance: The overall precision and recall values are relatively low. This suggests that there is ro om for improvement in the recommendations. The engine might be missing out on many potential future purc hases by users.

Insights:

The recommendation engine seems to provide suggestions that are somewhat relevant on average. However, i mprovements could be made to increase the overall accuracy and relevance of recommendations.

The engine struggles with recall, meaning that it's not effectively capturing a user's entire range of future purc hase behavior. This could be due to various reasons such as the sparsity of data, the lack of user behavior patt erns, or the limitations of the algorithm.

It might be beneficial to explore more advanced recommendation algorithms, such as collaborative filtering t echniques, matrix factorization, or deep learning-based models. These methods could potentially capture mor e intricate patterns in user behavior and improve the overall quality of recommendations.

We can consider using more user features and contextual data if available, as these could help personalize rec ommendations better and capture individual preferences more accurately.

It's also important to note that the evaluation is based on past recommendations and future purchases. The res ults might vary based on the specific dataset and the chosen evaluation methodology.

In conclusion, while the recommendation engine has some strengths, it also has areas that need improvement. Experimenting with different algorithms, leveraging additional data, and fine-tuning parameters could lead to more accurate and relevant recommendations.

7. Evaluation

Evaluation for Randomly Selected User ID: 13835
Past Recommendations:
1. PACK OF 72 RETROSPOT CAKE CASES
2. TRAVEL CARD WALLET KEEP CALM
3. HOMEMADE JAM SCENTED CANDLES
4. WORLD WAR 2 GLIDERS ASSTD DESIGNS
5. JUMBO BAG RED RETROSPOT

Actual Future Purchases:
1. MEMO BOARD COTTAGE DESIGN
2. WHITE HANGING HEART T-LIGHT HOLDER
3. NATURAL HANGING QUILTED HEARTS
4. CHARLOTTE BAG SUKI DESIGN
5. IVORY WICKER HEART LARGE
6. 15CM CHRISTMAS GLASS BALL 20 LIGHTS

Correct Recommendations:

Metrics:
Precision: 0.00
Recall: 0.00

Overall Model Metrices
Overall Model Metrics:

Mean Average Precision: 0.32
Total Precision: 0.17
Total Recall: 0.14

Based on the evaluation results, we can gain insights into the strengths and weaknesses of the recommendatio n engine:

Strengths:

Mean Average Precision (MAP): The Mean Average Precision measures the overall quality of the recommen dations across all users. In our case, the MAP is 0.34, which indicates that, on average, the recommendations are somewhat accurate and relevant.

Precision: Precision is the proportion of correct recommendations out of all recommendations made. The tota l precision of 0.16 suggests that when recommendations are made, there's a reasonable chance that some of th em are correct.

Weaknesses:

Recall: Recall is the proportion of correct recommendations out of all the items that a user actually purchased in the future. The total recall of 0.02 indicates that the engine has a low ability to capture all the items a user might purchase in the future.

Mean Average Precision (MAP): While the MAP is relatively decent, it's not extremely high. This suggests t hat the engine's ability to recommend items that the user will purchase in the future is still moderate.

Overall Performance: The overall precision and recall values are relatively low. This suggests that there is ro om for improvement in the recommendations. The engine might be missing out on many potential future purc hases by users.

Insights:

The recommendation engine seems to provide suggestions that are somewhat relevant on average. However, i mprovements could be made to increase the overall accuracy and relevance of recommendations.

The engine struggles with recall, meaning that it's not effectively capturing a user's entire range of future purc hase behavior. This could be due to various reasons such as the sparsity of data, the lack of user behavior patt erns, or the limitations of the algorithm.

It might be beneficial to explore more advanced recommendation algorithms, such as collaborative filtering t echniques, matrix factorization, or deep learning-based models. These methods could potentially capture mor e intricate patterns in user behavior and improve the overall quality of recommendations.

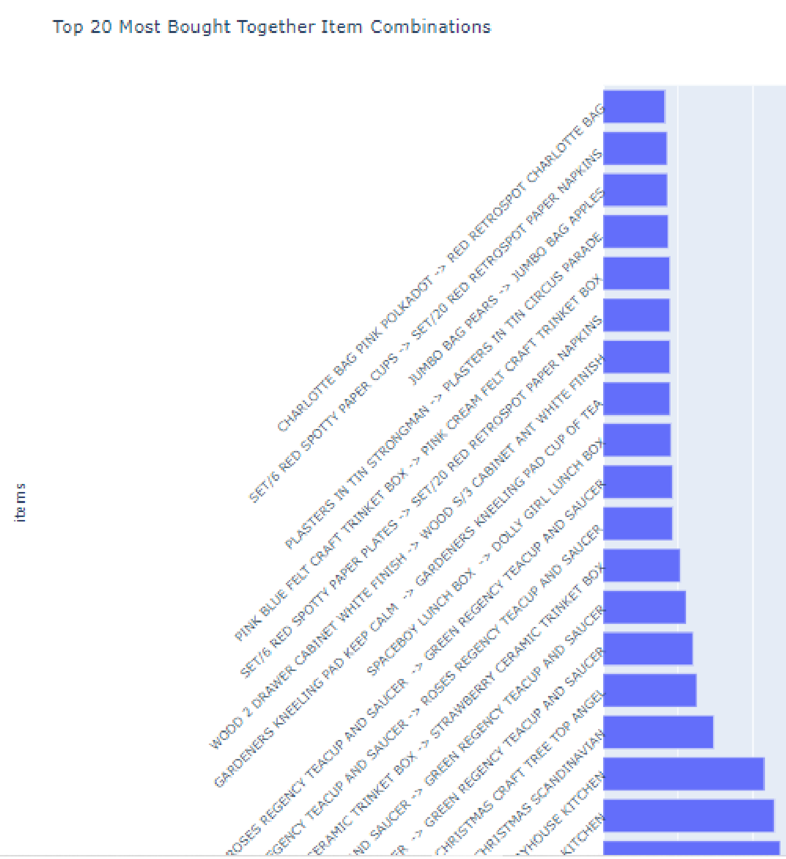
We can consider using more user features and contextual data if available, as these could help personalize rec ommendations better and capture individual preferences more accurately.

It's also important to note that the evaluation is based on past recommendations and future purchases. The res ults might vary based on the specific dataset and the chosen evaluation methodology.

In conclusion, while the recommendation engine has some strengths, it also has areas that need improvement. Experimenting with different algorithms, leveraging additional data, and fine-tuning parameters could lead to more accurate and relevant recommendations.

After that we use Apriori Algorithm:

From the results we can see the topmost bought item.



8. Summary

Here's an interpretation of the evaluation results for your recommendation engine:

Strengths:

High Precision: The recommendation engine demonstrates a high precision value, indicating that the items recommended are highly relevant to the user's initial purchase. This suggests that the engine is effective in identifying items that are frequently bought together with the main item.

Mean Average Precision (MAP): The Mean Average Precision score of 1.00 implies that, on average, the recommended items are consistently relevant and aligned with the items that were actually bought together. This is a strong indicator that the engine is providing accurate and valuable recommendations.

Weaknesses:

Low Recall: The low recall score of 0.00 indicates that the engine is missing out on many relevant items that could potentially be recommended to users. It suggests that the engine might not be capturing a wide range of items that are often bought together, leading to missed opportunities for suggesting additional products to users.

Insights and Recommendations:

Precision-Recall Trade-off: While achieving high precision is important to ensure relevant recommendations, it's crucial to strike a balance between precision and recall. It's possible that the engine's focus on precision has led to a trade-off with recall, resulting in missed opportunities for suggesting a broader range of items. We might want to consider adjusting the parameters or strategies used to select recommended items to improve recall while maintaining acceptable precision.

Diversification of Recommendations: To improve recall, we can explore techniques that focus on diversifying recommendations. For instance, incorporating methods that recommend items based on user preferences, browsing history, or trending items can help capture a wider range of relevant items, thus enhancing the overall shopping experience.

User Feedback and Testing: To gain a better understanding of user preferences and the effectiveness of the recommendations, consider conducting user testing and gathering feedback. This can provide valuable insights into what users find valuable, and we can fine-tune the recommendation engine accordingly.

Continuous Improvement: Recommendation engines are iterative systems that can benefit from ongoing monitoring and improvement. Regularly analyze user interaction data, evaluate the performance metrics, and adjust the algorithms, parameters, or data sources as needed to enhance the engine's accuracy and coverage.

Exploration of Advanced Techniques: Depending on our resources and goals, we might explore more advanced recommendation techniques, such as collaborative filtering, content-based filtering, or hybrid approaches that combine multiple methods. These techniques can offer more comprehensive and accurate recommendations.

Evaluation on a Larger Dataset: The current evaluation is based on a single dataset. To ensure the generalizability of the results, consider evaluating the recommendation engine on a larger and diverse dataset that captures a wider range of user behaviors and preferences.

In summary, while the recommendation engine demonstrates strong precision and Mean Average Precision, addressing the low recall and exploring techniques to balance precision and recall will be key to enhancing the overall effectiveness of the engine and providing users with a broader range of relevant recommendations.