USER PC

[Email address]

Abstract

[Draw your reader in with an engaging abstract. It is typically a short summary of the document.   
When you’re ready to add your content, just click here and start typing.]

[Document title]

[Document subtitle]

Title: Statistical Programming Assessment Report

Module Code:

Student ID:

Submission Date:

Word Count:

Contents

[Introduction 3](#_Toc182317587)

[Exploratory Data Analysis 3](#_Toc182317588)

[2.1 Monthly Fluctuation in Revenue and Transactions 3](#_Toc182317589)

[Monthly Revenue Trend 4](#_Toc182317590)

[Key Insights from the Monthly Revenue Trend: 5](#_Toc182317591)

[Key Insight from the Monthly Transaction Data 7](#_Toc182317592)

[Key Insights from the trend chart 9](#_Toc182317593)

[2.3 Seasonal Sales Variation 10](#_Toc182317594)

# Introduction

The purpose of this report is to perform an exploratory data analysis (EDA) on historical transaction data for an e-commerce company. The company seeks to optimize its sales and marketing strategies by gaining deeper insights into customer purchasing behaviour, product performance, and revenue trends.

The dataset contains key variables such as customer ID, transaction date, product ID, product category, quantity purchased, and total price. By analyzing this data, we aim to uncover patterns and trends that can guide the company in making data-driven decisions.

To achieve this, Python was used as the primary tool, leveraging powerful libraries such as Pandas for data manipulation and Matplotlib/Seaborn for visualizations. The analysis focuses on four key objectives:

1. Analysing monthly fluctuations in total revenue and transaction count.
2. Identifying product categories with the highest revenue and consistent growth.
3. Exploring seasonal variations in product sales.
4. Examining shifts in customer purchasing behaviour across multiple transactions.

This report will present actionable insights supported by visualizations and tables to help the company enhance its sales strategy effectively.

# 

# Exploratory Data Analysis

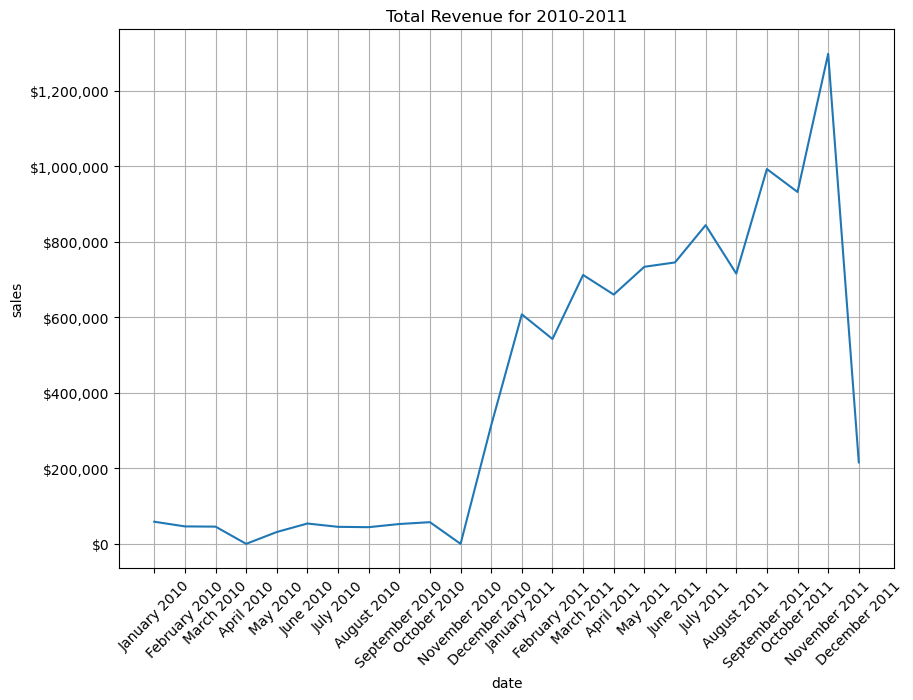
## 2.1 Monthly Fluctuation in Revenue and Transactions

#### **Data Preparation**

The first step in analyzing monthly fluctuations in revenue was to calculate total sales for each transaction. This was achieved by introducing a new column, Sales, computed as the product of Quantity and UnitPrice. The dataset was then aggregated on a monthly basis using Python’s Pandas library to sum the revenue for each month.

|  |  |  |
| --- | --- | --- |
| **index** | **Month** | **Revenue($)** |
| 1. | January, 2010 | 58635.560 |
| 2. | February, 2010 | 46207.280 |
| 3. | March, 2010 | 45620.460 |
| 4. | April, 2010 | 0.000 |
| 5. | May, 2010 | 31383.950 |
| 6. | June, 2010 | 53860.180 |
| 7. | July, 2010 | 45059.050 |
| 8. | August, 2010 | 44189.840 |
| 9. | September, 2010 | 52532.130 |
| 10. | October, 2010 | 57404.910 |
| 11. | November, 2010 | 0.00 |
| 12. | December, 2010 | 314063.660 |
| 13. | January, 2011 | 607748.110 |
| 14. | February, 2011 | 542511.540 |
| 15. | March, 2011 | 712059.510 |
| 16. | April, 2011 | 660149.571 |
| 17. | May, 2011 | 733769.320 |
| 18. | June, 2011 | 745252.340 |
| 19. | July, 2011 | 843891.231 |
| 20. | August, 2011 | 715632.720 |
| 21. | September, 2011 | 992555.452 |
| 22. | October, 2011 | 931852.160 |
| 23 | November, 2011 | 1297606.420 |
| 24. | December, 2011 | 215762.540 |

## **Monthly Revenue Trend**



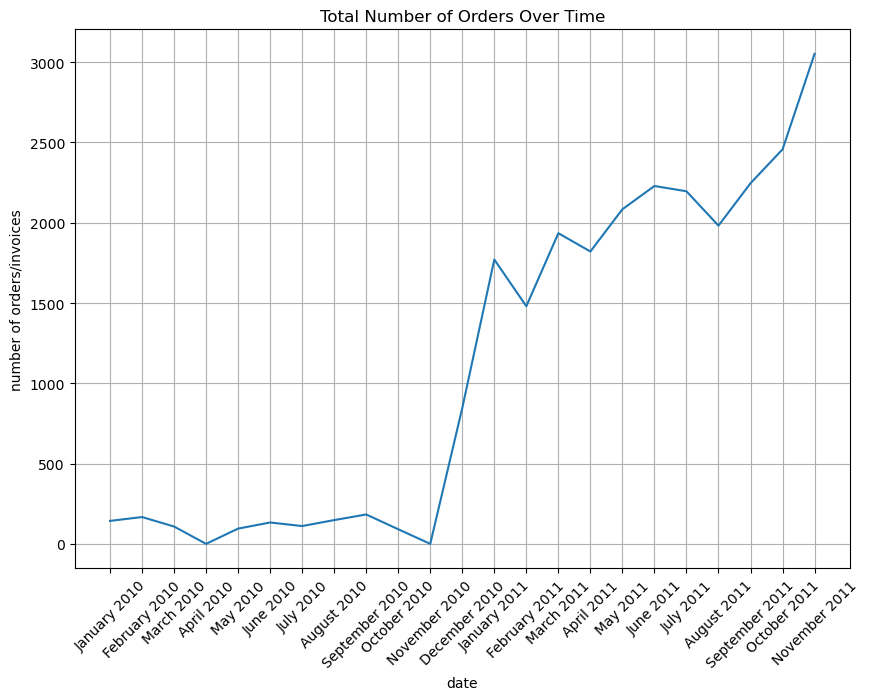
### Key Insights from the Monthly Revenue Trend:

1. **Steady Growth in Revenue**:
   * There is a noticeable upward trend in revenue, particularly from mid-2010 to late 2011. This indicates a period of business growth and increased customer activity.
2. **Seasonal Peaks**:
   * Revenue spikes are observed around November and December, suggesting that the business may benefit from seasonal events such as holiday shopping or end-of-year sales.
3. **Sudden Drop in December 2011**:
   * Despite the usual increase in revenue during the last quarter, December 2011 shows a sharp decline. This could indicate a data anomaly, an operational issue, or external factors such as economic downturns. According analysis this is due to the incomplete data recorded. Data spans from December 01 – 10, 2010
4. **Low Revenue Periods**:
   * The early months in the dataset (January to June 2010) show relatively low revenue, possibly indicating that the business was either in a nascent stage or undergoing challenges during that period.
5. **Opportunities for Analysis**:
   * The sharp increase from August 2010 could be linked to successful marketing campaigns, product launches, or expansion. Understanding what drove this growth can help replicate the success in the future.

Monthly Transaction Trend

|  |  |  |
| --- | --- | --- |
| **index** | **Month** | **Number of Orders** |
| 1. | January, 2010 | 143 |
| 2. | February, 2010 | 167 |
| 3. | March, 2010 | 108 |
| 4. | April, 2010 | 0 |
| 5. | May, 2010 | 95 |
| 6. | June, 2010 | 133 |
| 7. | July, 2010 | 111 |
| 8. | August, 2010 | 148 |
| 9. | September, 2010 | 183 |
| 10. | October, 2010 | 91 |
| 11. | November, 2010 | 0 |
| 12. | December, 2010 | 846 |
| 13. | January, 2011 | 1771 |
| 14. | February, 2011 | 1480 |
| 15. | March, 2011 | 1935 |
| 16. | April, 2011 | 1821 |
| 17. | May, 2011 | 2084 |
| 18. | June, 2011 | 2229 |
| 19. | July, 2011 | 2196 |
| 20. | August, 2011 | 1982 |
| 21. | September, 2011 | 2246 |
| 22. | October, 2011 | 2457 |
| 23 | November, 2011 | 3057 |
| 24. | December, 2011 | -------- |

Note: December, 2011 will be misrepresented with an incomplete data so it was disregarded.



### Key Insight from the Monthly Transaction Data

1. There was a dramatic increase in orders starting from October-November 2010, showing a clear inflection point in the business. Before this point, orders were relatively flat and low (around 200 orders per month).

2. After the sharp increase in late 2010, the business maintained a generally upward trend through 2011, with order volumes consistently above 1,500 units.

3. The most recent period (late 2011) shows particularly strong growth, reaching approximately 3,000 orders - the highest point in the entire timeline.

4. There are some seasonal fluctuations visible in 2011, with minor dips and peaks, but the overall trend remains positive.

5. The jump from around 200 orders to over 1,500 orders represents roughly a 750% increase in business volume, suggesting a major business event occurred (possibly a new marketing strategy, product launch, or market expansion).

The most significant insight is the transformational change that occurred in late 2010, which marked a permanent shift from a small-volume operation to a much larger-scale business, followed by sustained growth

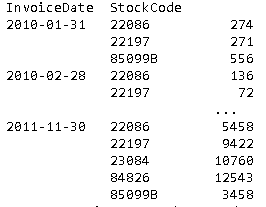
Product Category Revenue Analysis

The table below show the top-performing based on total revenue over the time

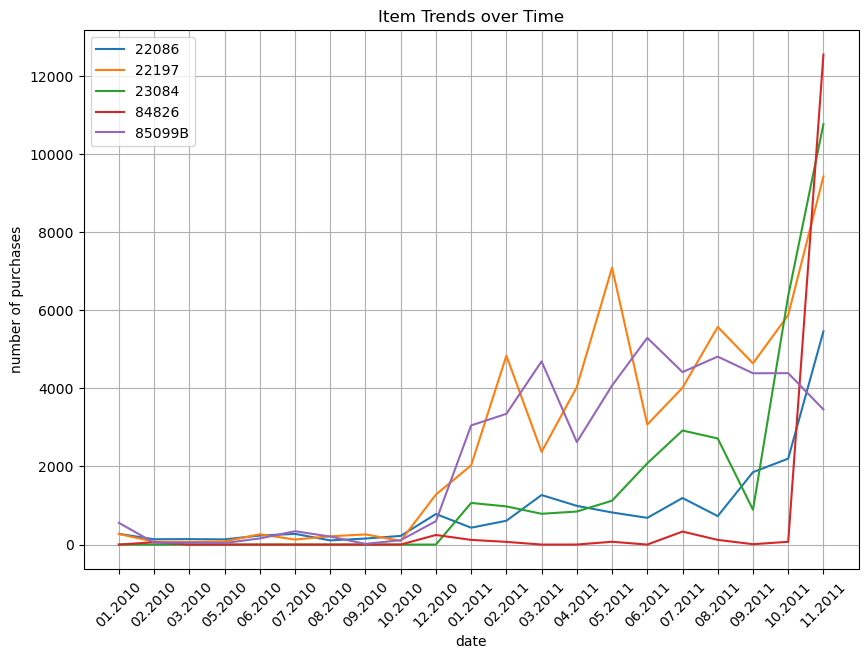
|  |  |  |
| --- | --- | --- |
| Index | StockCode | Quanity |
| 1 | 84826 | 12543 |
| 2 | 23084 | 10760 |
| 3 | 22197 | 9422 |
| 4 | 85123A | 6842 |
| 5 | 22086 | 5458 |
| 6 | 84879 | 5129 |
| 7 | 84077 | 4699 |
| 8 | 21108 | 4541 |
| 9 | 22578 | 3677 |
| 10 | 22577 | 3640 |

As we can see from the result, the products with codes 84826, 23084, 22797, 85123A and 22086 were the top best sellers in the month November 2011.

Now that we know what the top 5 sellers were in November 20, lets aggregate the monthly sales data for these five products.



The figure below shows the trend.



### Key Insights from the trend chart

1. Major Surge in Late 2011:

- Multiple items (particularly 84826 and 23084) show a dramatic spike in purchases in November 2011

- This surge is far above previous levels, reaching around 12,000 units for item 84826

2. Different Growth Patterns:

- Item 22197 (orange line) shows more volatility throughout 2011 with several peaks and troughs

- Item 85099B (purple line) shows moderate but consistent performance through most of 2011

- Item 22086 (blue line) maintains the most stable pattern with gradual growth

- Item 84826 (red line) remains relatively flat until the dramatic increase at the end

3. Timing of Changes:

- Similar to the previous graph, there's a notable change in behaviour starting around late 2010/early 2011

- Before this point, all items had relatively low and stable purchase numbers

- After this point, there's increased volatility and generally higher volumes

4. Performance Comparison:

- Items showed different levels of success:

- Some maintained steady growth

- Others showed high volatility

- A few remained relatively flat until the final surge

5. The most striking insight is the synchronized sharp uptick in multiple items at the end of the period, suggesting a possible seasonal effect or company-wide event affecting multiple product lines simultaneously.

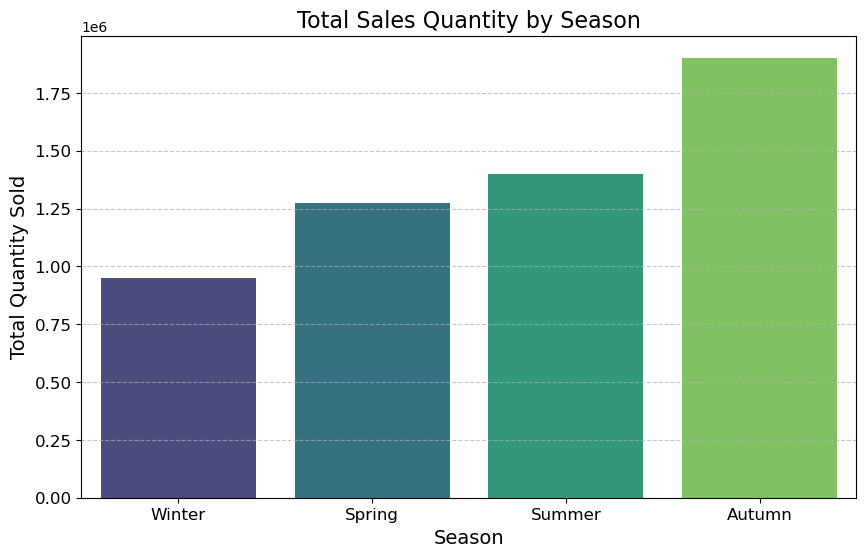
## 

## 2.3 Seasonal Sales Variation

The table below shows the product category variation across different seasons analysing the top performing products during each seasons ranging from summer, autumn, spring and winter.

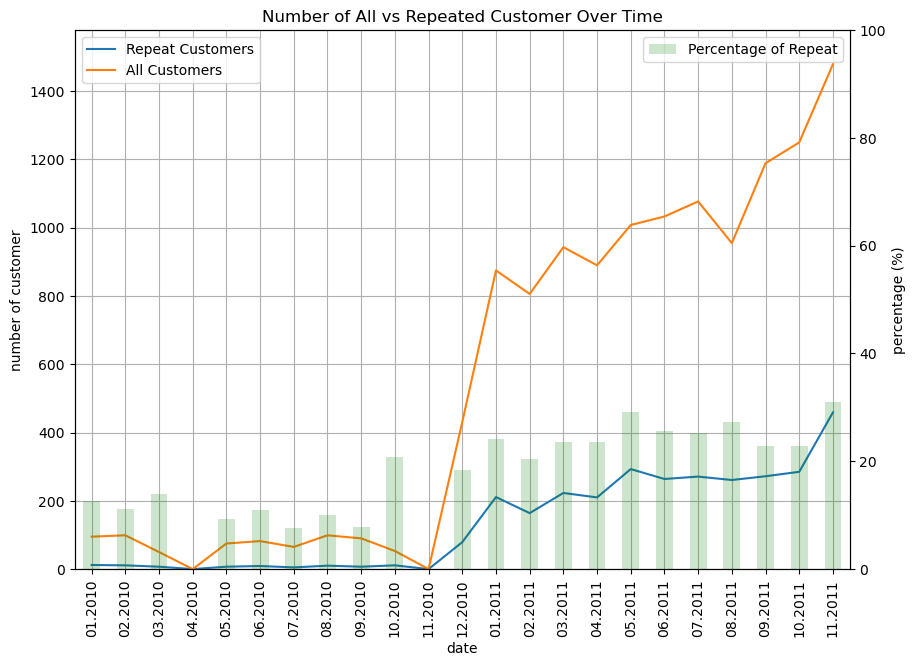
|  |  |  |
| --- | --- | --- |
| Season | Description | Quantity Sold |
| Autumn | Paper craft, little birdie | 80995 |
| Winter | Medium ceramic top storage jar | 74327 |
| Autumn | Popcorn holder | 18715 |
| Spring | World war 2 Glider assisted designs | 18452 |
| Autumn | Rabbit Night Light | 18048 |
| Autumn | World war 2 gliders assisted design | 16053 |
| Summer | Jumbo bag red retrospot | 15554 |
| Autumn | Jumbo bag red retrospot | 12990 |
| Autumn | Assisted design 3D paper stickers | 12624 |
| Spring | Assorted Colours Silk Fan | 12311 |

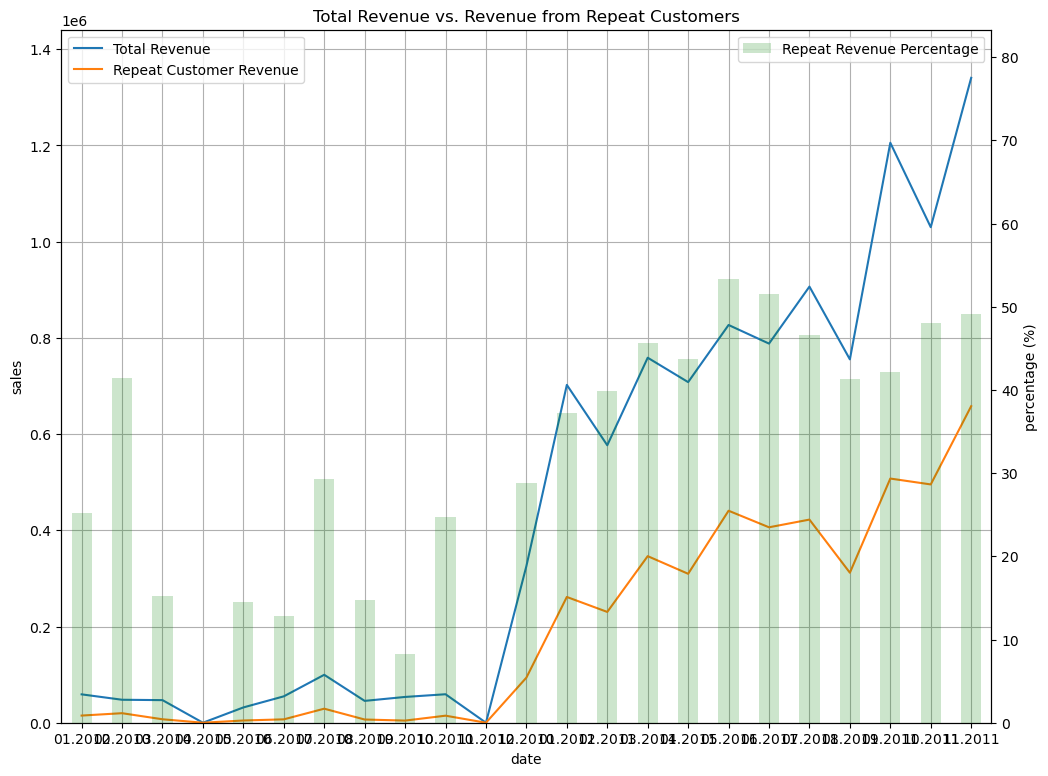
The total Sales Quantity by Season



Generally, we have more sales during Autumn.

## 2.4 Customer Purchasing behaviour

Fig 1. 



focusing on the 2010 to 2011 time period in the image:

Valuable Insights:

1. The total revenue increased significantly from 2010 to 2011, going from around 0.8 to over 1.4.
2. The revenue from repeat customers also grew rapidly during this period, going from around 0.2 to over 0.4.
3. The repeat customer revenue percentage increased from around 25% in 2010 to over 60% in 2011.

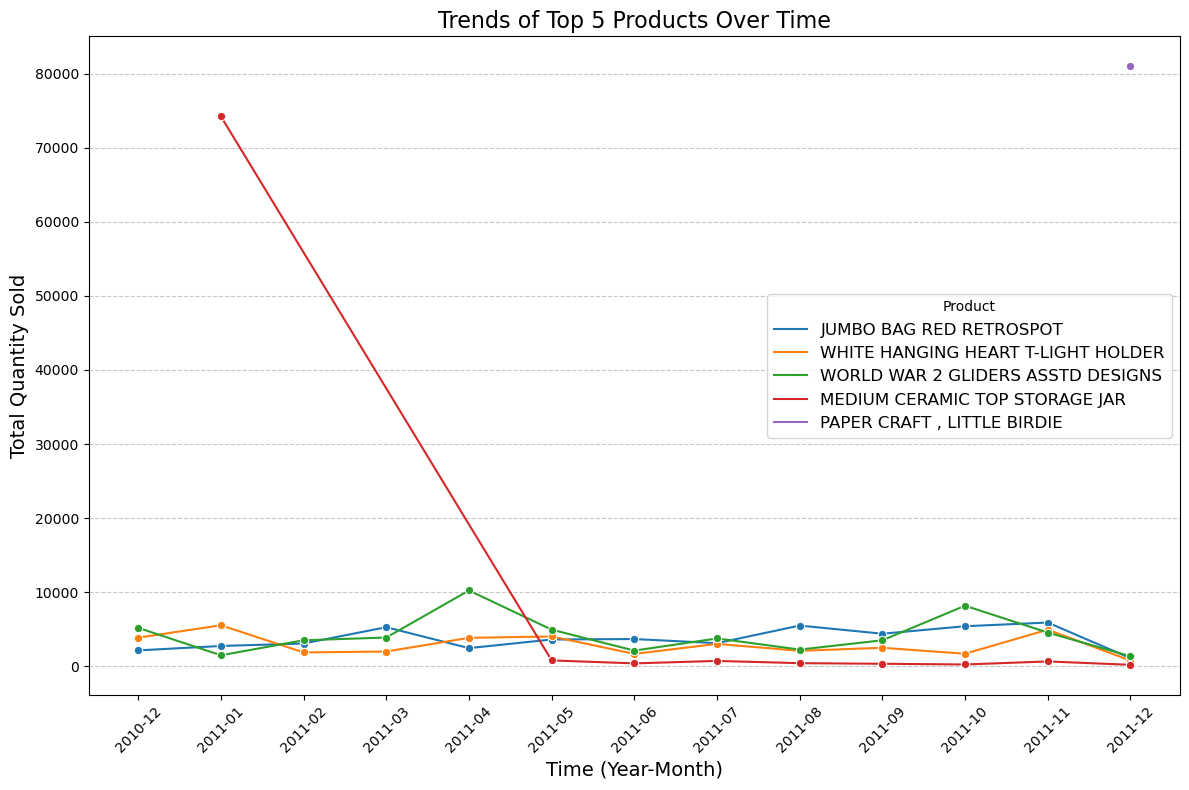
Action Insights:

1. The company should continue focusing on retaining and growing its repeat customer base, as this segment is driving a larger portion of total revenue.
2. Analyse what strategies or initiatives led to the sharp increase in repeat customer revenue from 2010 to 2011, and look for ways to further capitalize on these successful approaches.
3. Consider investing more resources into customer loyalty programs, repeat purchase incentives, or other initiatives that can help maintain and build upon the strong repeat customer growth seen in this time period.
4. Evaluate the overall customer acquisition and retention strategies to understand what is driving the total revenue growth, and see if there are opportunities to further optimize these efforts.

We went further to discover the top customers as well as the top 5 product overtime

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 13089.0  index | CustomerID | AvgTransaction Value | Avg Items Per Purchase | Unique Product Per Transaction | Purchase frequecy |
| 1. | 12748.0 | 160.570143 | 122.609524 | 20.400000 | 210 |
| 2. | 14911.0 | 715.547562 | 400.572139 | 28.218905 | 201 |
| 3. | 17841.0 | 330.577177 | 186.056452 | 60.282258 | 124 |
| 4. | 13089.0 | 606.451856 | 320.309278 | 18.680412 | 97 |
| 5. | 14606.0 | 130.716667 | 66.924731 | 28.602151 | 93 |
| 6. | 15311.0 | 667.779121 | 419.714286 | 25.901099 | 91 |
| 7. | 12971.0 | 130.115233 | 108.011628 | 3.546512 | 86 |
| 8. | 14646.0 | 3786.567838 | 2668.797297 | 28.108108 | 74 |
| 9. | 16029.0 | 1286.108571 | 638.222222 | 3.809524 | 63 |
| 10. | 13408.0 | 453.500645 | 261.806452 | 7.709677 | 62 |

.



Analysing the spending pattern of each top customer

