**Postgraduate Diploma in Supply Chain Analytics – Data Analytics + Visualisation**

Lecturer: Shagufta Henna

Student: Andrew Shanahan

Student ID: L00187317

**Project found on GitHub:** [**https://github.com/AndrewShanahan/SupplyChain/tree/main**](https://github.com/AndrewShanahan/SupplyChain/tree/main)

*Please note that purpose of GitHub repository is continue to work on this project in order to create a comprehensive project for professional purposes and therefore will remain a work in progress following completions of this module/course*

**Retail Transactions Dataset - Report**

**Executive Summary**

This report presents the findings of a Supply Chain Data Analytics project conducted as part of the Postgraduate Diploma in Supply Chain Analytics course from Atlantic Technological University. This project has leveraged data analytics methods to gain insights into a selected dataset and identify opportunities to improve efficiencies and to optimise operations.

**Introduction**

This project has been conducted to analyse a retail dataset in the context of customer satisfaction related to the delivery of the purchased goods. My objectives with this project were to perform some data analysis of a Supply Chain related dataset to identify key insights. This analysis was using Python in Jupyter Notebooks. Most of the initial analysis focused on a statistical analysis approach and from there I moved onto a more in-dept look and then some visualisations. I also created an interactive dashboard using Power BI and designed to emulate a self-service style dashboard.

**Dataset Description**

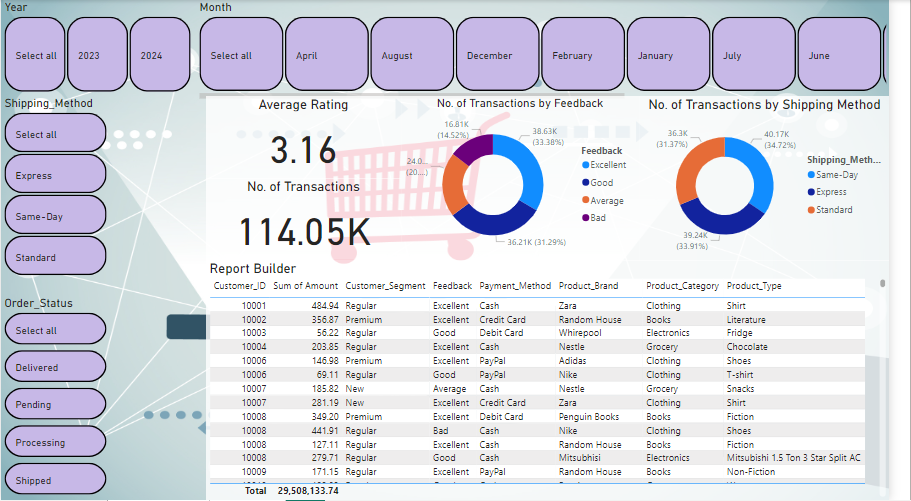
The dataset used in this project was a Retail Transactional Dataset sourced from Kaggle [1]. The file is in a CSV format and is quite a large dataset therefore this was useful for drawing accurate conclusions regarding the statistical analysis.

**Methods Section**

* Data Collection: The dataset used in this project was sourced from Kaggle.com.
* Data Analysis: Python and Jupyter notebooks were utilised for the data cleaning, exploration and analysis. A number of libraries were used, some key ones included: pandas, NumPy, seaborn and matplotlib.
* Dashboard Creation: An interactive dashboard was created using Power BI and this showcased the ability to present data in such a way for varied audiences, in this case, I built this in mind of multiple users from different areas of expertise that may be using the dashboard and therefore focused on making it a self-service style where they can select a number of filters/slicers to filter to specific interests.

The screenshots below are images of the how the dashboard is presented. The first image is the Home or landing page where it allows users to navigate to a hypothetical functional spec document where it would have details on the build in addition to file owner and deputy owner. It also has hypothetical link to how a user could add/remove or amend their access. This allows for different users to have different levels of access. For example, a HR Dashboard may have different levels of access depending on the users’ needs. The second image is a screenshot of how the dashboard would look. I followed a principle that I would have used in one of my previous roles a QlikView developer where the top of the page and left-hand side of the page were dedicated to calendars and filtering options. The middle of the screen dedicated to some important KPI’s, statistics or graphs and the bottom of the screen a report builder type option, this allows for a better user experience as they can tailor their own use of the dashboard.2

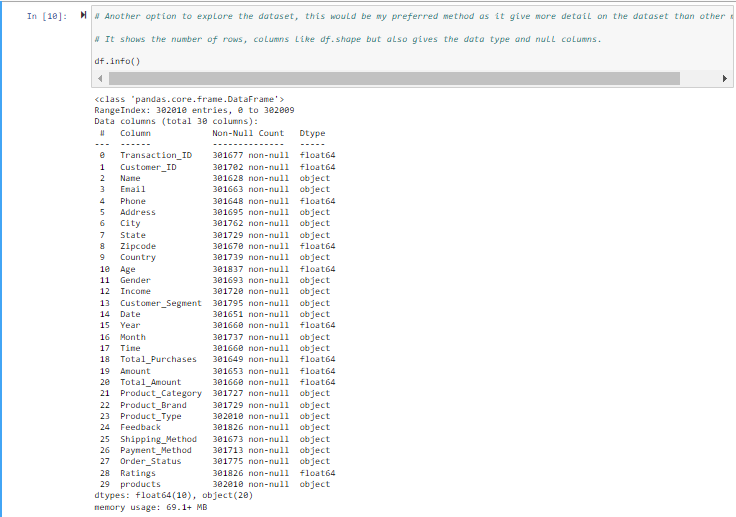
****

****

**Results**

Initial analysis of the dataset had a look at the how the dataset would look by looking at the first and last 5 rows. Next, I analysed the summary statistics of the dataset where I got the count, mean, standard deviation, min, 25%, 50%, 75% and max. I believe this is a great was of analysing the data especially for analysing the distribution of the data. Further analysis of the dataset checked for any null values, checked the data types, checked the size of the dataset and had a look at the columns.

Another useful check on the data is checking the information which gives a very comprehensive result as shown in the screenshot below;



The next step was Data cleaning, where the date column had to be amended to datetime as the data type was not recognised. Some further amendments were made to other columns in order to correctly analyse these columns. Further work was required on the month columns as per screenshot below;







The next step of the analysis was creating a number of visuals to showcase ability to plot using python. While the dataset contains data for 2023 and 2024, I thought it would be an interesting insight to see what months are most popular for sales which could be very useful for making strategic decisions.



Another visual I wanted to create was to monitor sales by years, it would be more useful with a larger dataset ranging over a longer number of years but it was still something that can demonstrate the potential of using visuals in python.



**Conclusion**

This project has leveraged data analytics methodologies to extract valuable insights from the dataset. Through a combination of Python-based statistical analysis and visualizations, as well as the creation of an interactive dashboard using Power BI, this project demonstrates the potential of data analytics to enhance decision-making in supply chain management.

The initial data analysis in Python allowed for a thorough examination of the dataset, revealing key trends and patterns that could impact customer satisfaction and operational efficiency. The summary statistics provided a foundational understanding of the dataset's distribution, while the identification and treatment of null values and data type corrections ensured the integrity and accuracy of the subsequent analyses.

The visualization of sales data by month and year highlighted important trends, offering strategic insights that can inform future sales strategies and inventory management decisions. These visualizations not only underscore the importance of data-driven decision-making but also demonstrate the efficacy of Python as a tool for data analysis and visualization.

The development of a self-service interactive dashboard in Power BI further illustrates the versatility and accessibility of data analytics tools. By enabling users from various functional areas to filter and interact with the data according to their specific needs, the dashboard promotes a more inclusive and informed decision-making process across the organization.

**Future Work**

This project lays the groundwork for ongoing analysis and optimisation in supply chain management. Some elements to consider for future work could be:

* Expanding the Dataset: Incorporating multiple years to provide more comprehensive insights and trends.
* Machine Learning: Advanced Predictive Analytics, applying machine learning models to predict future sales and demand, further enhancing inventory management and operational efficiency, this is something I would have liked to include but time was an obstacle and will be something that I add to this project at a later date on my GitHub repository.
* Real-Time Data Integration: In a real-world environment, integrating real-time data feeds into the Power BI dashboard to enable dynamic, up-to-date decision-making.
* User Access Controls: Developing more sophisticated user access controls to ensure data security and privacy while promoting a collaborative environment, this is something that gets more complicated to the more effort that goes into it but allows for a better tailored experience for the user as it will and should only give the user certain access based on their role.
* My GitHub repository for this project will remain active, reflecting enhancements and expansions to create a comprehensive, professional-grade supply chain analytics tool.

**References**

[1] <https://www.kaggle.com/datasets/bhavikjikadara/retail-transactional-dataset?resource=download>