# Student ID: 4214293

Discrete maths coursework

CONSULTANCYREPORT

Contents

[Student ID: 4214293 1](#_Toc153393329)

[Coursework aim: 3](#_Toc153393330)

[Consultancy report 3](#_Toc153393331)

[Uploading the csv file: 3](#_Toc153393332)

[Loading the csv file: 3](#_Toc153393333)

[Data processing: 4](#_Toc153393334)

[Different types of calculated data 5](#_Toc153393335)

[*Overview of the dataset above:* 5](#_Toc153393336)

[*Customer demographics:* 5](#_Toc153393337)

[*Discounts* 5](#_Toc153393338)

[Analysis of sales dynamics and customer insights 7](#_Toc153393339)

[Data analysis: 7](#_Toc153393340)

[Product analysis: 7](#_Toc153393341)

[Discount impact on sales analysis: 8](#_Toc153393342)

[Geographic insights: 10](#_Toc153393343)

[Customer demographics: 12](#_Toc153393344)

[Conclusion 14](#_Toc153393345)

[Reference list 15](#_Toc153393346)

# Coursework aim:

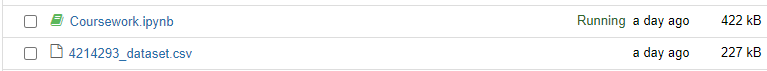
The scenario is as follows. Imagine that you are freelancing data analyst providing data information on different datasets provided to you. You will provide insights such as on sale of a products, quantity of products based on discounts, product status based on price. The more insights you can give them, based purely on the data, the better.

# Consultancy report

In this case, I am a data analyst and will analyse the different types of data and what types of correlations and representations it has with the data. With the help of python libraries. For this report, I will be using Jupyter notebook as this will allow me to record my data and code into this document. I will be taking screenshots of my code and my diagrams (I.e. different types of graphs and tables with data in them).

## Uploading the csv file:

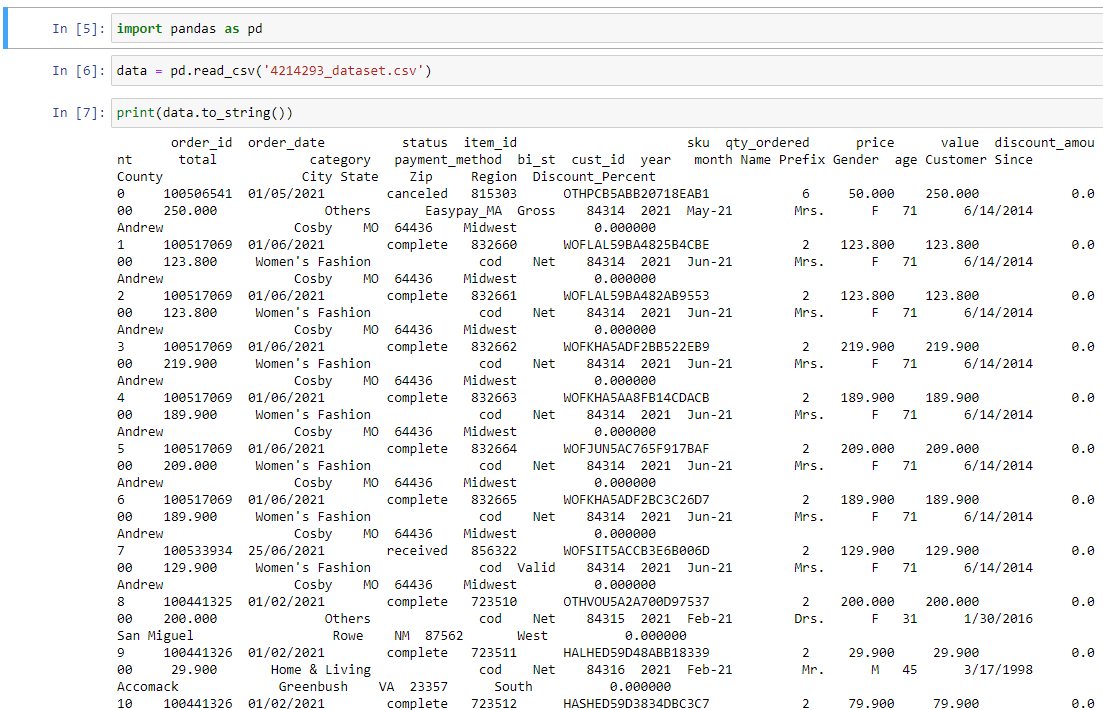
I need to import the data as a csv file to my Jupyter folder, it contains all the data which is needed for my data analysis:



After that, I will create a new ipython file (also known as ipykernal) and I will call it, Coursework (as seen above).

### Loading the csv file:

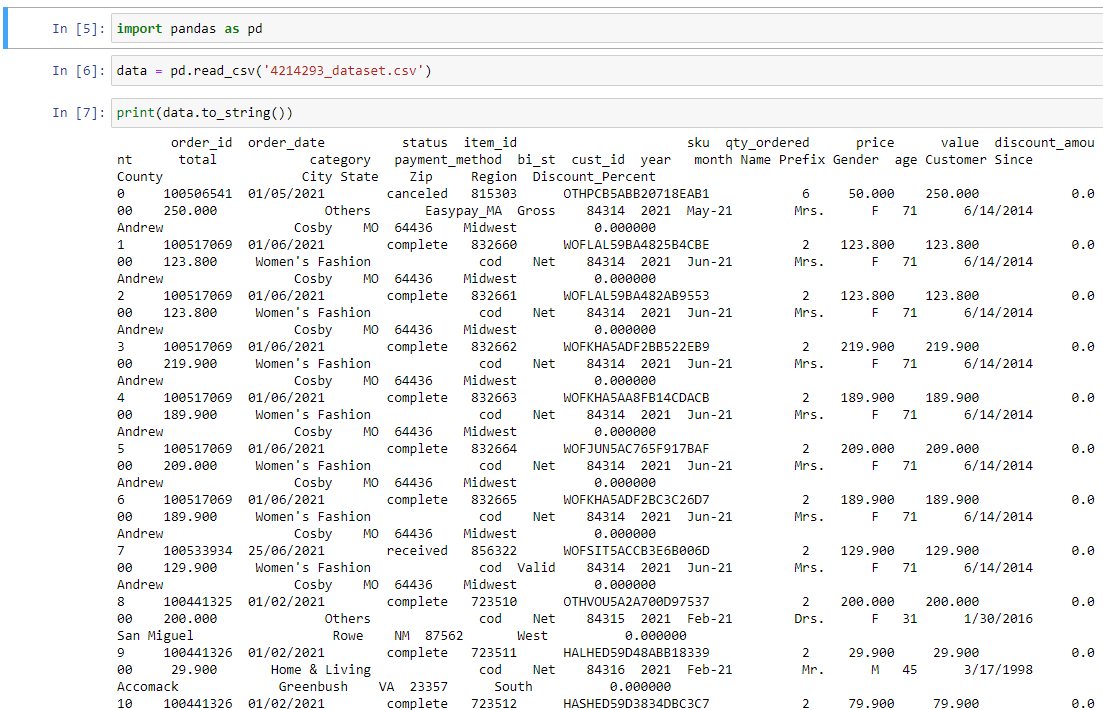
As I will be using a csv file to import the data into the program to analyse and create graphs. I will have to use a library called pandas as this will allow me to import the csv file.



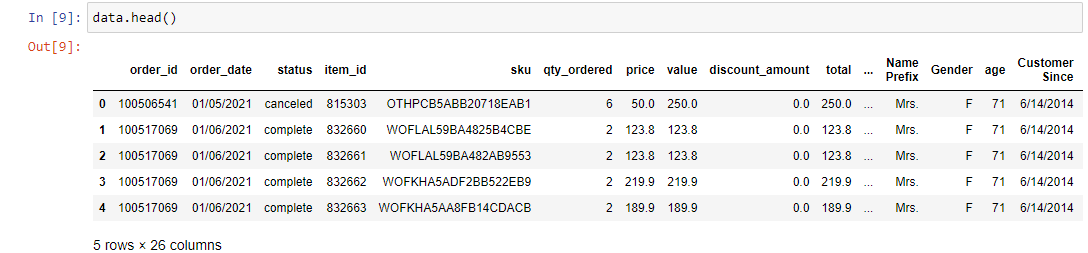
The dataset has been fully loaded into the jupyter file.

### Data processing:

CSV file output of what it looks like



After looking at this csv file in this format I will need to clean out the data. Which can be done by using the. head() function in python via the pandas library:



As seen above this shows what the first 5 rows of the dataset.

If I would like to see the last 5 rows of the dataset, I can use the. tail() function in python via the pandas library:

A screenshot of a computer

Description automatically generatedThis shows the last 5 rows of the dataset above.

## Different types of calculated data

1. Retrieving different types of calculated data:

Within my dataset I would like to retrieve data such as percentile, mean and std of the numerical values of the dataset I am working with. To do this, I will be using .describe() function from the panda’s library.

A screenshot of a computer

Description automatically generated

A screenshot of a calculator

Description automatically generated

1. Analysing the dataset above (with different calculations involved)

### *Overview of the dataset above:*

* By analysing this data above, it clear that it contains up to 1,219 records of E-Commerce sales transactions.
* The average of the quantity ordered is around 1.98 (to 2 decimal places).
* The average price is around $831.80, which brings the total value to $495.94 (to 2 decimal places).

### *Customer demographics:*

* The average for the customers according to the dataset is, 46.58 years.
* According to the dataset above, the average transactions were taken from the year 2021.
* The age of customers ranges from approximately 17-75 years old.

#### Discounts:

* The average discount amount is 31.19% per order (to 2 decimal places).
* The average discount percent was 3.74% (to 1 decimal place).
* The max discount percent was 63.03%, (to 2 decimal places), indicating that during times of the year such as Christmas holidays, boxing day shopping or maybe even black Friday sales the maximum discount percent went up to 63.03%.
* The maximum discount amount is 1000.
* The average discount 31.2 (to 1 decimal place).

#### Sales patterns:

* The average value of goods which were ordered is $4956.94 (to 2 decimal places)
* The maximum value of goods which were ordered is $10579.90 (to 2 decimal places).
* The minimum quantity of good that is ordered is 1.
* Whereas the maximum quantity of good that is ordered is 15.
* Total value is equal to $1219.

## Analysis of sales dynamics and customer insights

### Data analysis:

I have used:

* (www.w3schools.com, n.d.) – for creating and representing histogram, and for creating a bar chart. **Please note that the citation I have provided are the same but from I have taken information from different pages from the w3 schools website to help me create my graphs in python.**
* (GeeksforGeeks, 2021) – for creating and representing the pie chart.

### Product analysis:

A pie chart with numbers and a pie chart

Description automatically generatedBy finding out which category of products are popular; I will first need to analyse the dataset to give me an insight into which category of products are popular. To help me analyse this, I created a pie chart with the help of the matplot.lib as this allows me to see which categories the most popular and which categories are the least popular. By analysing this pie chart, mobiles and tablets are the most popular category of products to purchase by consumers which is about 28.3%. This is due to the increase in demand for those goods during 2021 as covid-19 forced people to work from home, and children to take online school. Hence, the rapid rise in the demand for mobiles and tablets allows people to have the most up-to-date technologies as well as being able to work and communicate remotely. Women’s fashion is the second most popular category type, since women enjoy shopping for clothes the demand online has risen. Hence one of the possibilities is that women’s fashion is highly popular. The least popular category type is books which falls under 1%, which is since you can find any book online in a PDF form without having to pay money for it or watch videos and read articles about a particular product. From this pie chart, the high profits come from selling mobiles and tablets as it has become a necessity for consumers to have, whether it be for work or entertainment purposes, and make work easier for people as they can work from anywhere in the work with the use of a laptop or tablet. Being able to have meetings on applications such as Google Meet, zoom, Microsoft teams etc. The lowest profits are made from selling books as they are near or less than 1% (in terms of popularity).

### Discount impact on sales analysis:

A screen shot of a computer code

Description automatically generated

A graph with green dots

Description automatically generated

For me to examine how discount amount impacts on quantity ordered. I would need to analyse the dataset in terms of checking the quantity ordered when discounts are applied. As the discount increases, does the quantity ordered increases? This is what I wanted to find out. What I have done for this investigation, is create a scatter graph, quantity ordered against discount which would allow to determine the overall correlation between the two variables. By looking at this scatter graph, as the as the discount amount rises the quantity orders of products also rises. This indicates that consumers will be more likely to purchase items more sooner, since they can get for cheaper, i.e., reduced price (lower than the original price). Therefore, the correlation between quantity ordered and discount is positive.

According to my research from (Volusion Ecommerce Blog | SMB Marketing, Design, & Strategy, 2023). One of the ways to optimise discount strategies, is by using strategies such as increasing sales over the holidays such as Christmas, and Black Friday which will lead to an increase in quantity ordered allowing the company to make profit and reinvest that profit into new product line or buying new stock for popular products. Another technique to increase sales is by doing discounts on new products which will encourage consumers to try their brand-new products and recommend it to their family and friends which may lead to increase in potential sales. Therefore, if sellers use these strategies to attract new customers and long-time customers, they can increase the demand (quantity ordered) for their products.

A screenshot of a computer screen

Description automatically generated

By analysing this histogram above, Where the quantity ordered between 1-7 as the quantity ordered rises these leads a fall in frequency which shows that consumers prefer to purchase good in small amounts like 1-4 quantity ordered of goods. it clear that a histogram would not be suitable for quantity ordered against frequency. This is because, there is an outlier between 11-12 quantity ordered.

### Geographic insights:

A pie chart with a pie chart and a diagram

Description automatically generated

For me to see which regions have the most sales, I need to see how many people purchase items from different regions such as South, West, Northeast, Midwest. Thus, I have created a pie chart which represents which regions are the highest performing in terms of sales. By analysing this pie chart, the south is the highest performing region with 33.8%, implying that demand for online shopping for products is higher, this is perhaps due to consumers preferring to shop online for their goods and services rather than in person. Because you can save time searching and paying for the product at checkout, whereas going to a physical store would mean that the consumer would have to be on the lookout for their product and having to ask staff for it and go through different aisle (sections) just to find a thing which they need. On the other hand, the northeast has the lowest out of all the three regions, about 13%. Which may show that consumers in that region may not find online shopping convenient and prefer to shop for their goods in person.

A screenshot of a computer program

Description automatically generated

A purple graph with black text

Description automatically generated

The state with the highest performance, in terms of sales is CA (California). According to my research (Inc (n.d.)), it is one of the places with rising economic growth and increase in business opportunities. Due to the growth of online business, technology and easy access to the internet has made it easier to setup businesses by reducing startup costs.

### Customer demographics:

A screenshot of a computer program

Description automatically generated

A graph of age versus frequency

Description automatically generated  
Since, I want to find out the age distribution of online shopping, I have created a histogram which shows a visual representation of ages of different consumers taken from my dataset I am analysing. Ages between 15 – 75 are sorted into 10 bins. People with ages between 70-75 have the highest frequency of about 170, and people with ages between 35-45 having the lowest frequency of around 70. This indicates that, people who are in the age of 70-75 use online shopping the most, which makes it convenient for them to get their stuff delivered to them to their front door, without having for them to get up and go outside to the store to get their necessities.

A screenshot of a computer program

Description automatically generated

A chart with green dots

Description automatically generated

Since I wanted to analyse which type of category types are popular amongst different age groups. I have coded a program that will generate me a scatter graph for the age against the category types. By analysing this scatter graph, ages between 15-30, 35-50, 65-75 have the most demand for mobiles and tablets and the category types with the lowest demand it between the range of 20-25 is school and education.

## Conclusion

Overall, I have conducted data analysis regarding the dataset that I was given. The key findings which I have found are: Mobile and tablets are the most popular category in terms of online shopping due to increased demand, as remote working becoming more favourable and online learning since schools, colleges and universities were closed due to Covid-19. There is a positive correlation between quantity ordered and discount amounts. As discounts increase, the quantity ordered increases.

California is the state with the highest demand for online shopping, entrepreneurial opportunities. The highest frequency for age group is 70-75 in terms of online shopping. Another thing to take into consideration, that people who live in the state of California live far away from their local shops, and having the ability to order things off sites such as Amazon makes it convenient for next day deliveries as there are multiple warehouses scattered across America.

A recommendation I would give for this specific e-commerce company, is that since their mobile and tablets are the most popular products amongst consumers, what they should do is apply discounts to laptops and tablets which are old and sell off the old stock to reinvest into new high-end tablets and laptops.

## Reference list

1. GeeksforGeeks. (2021). *How to Create Pie Chart from Pandas DataFrame?* [online] Available at: https://www.geeksforgeeks.org/how-to-create-pie-chart-from-pandas-dataframe/.
2. Inc, E. com (n.d.). *How to Start an Online Business in California*. [online] IncParadise. Available at: https://incparadise.net/california/start-online-business-in-california/#:~:text=California%20is%20considered%20as%20one [Accessed 3 Dec. 2023].
3. Volusion Ecommerce Blog | SMB Marketing, Design, & Strategy. (2023). *Leveraging the Psychology of Discounts to Make More Money | Volusion*. [online] Available at: https://www.volusion.com/blog/using-the-psychology-of-discounts-to-make-more-money/#:~:text=the%20digital%20space.- [Accessed 3 Dec. 2023].
4. www.w3schools.com. (n.d.). *Matplotlib Histograms*. [online] Available at: https://www.w3schools.com/python/matplotlib\_histograms.asp.
5. www.w3schools.com. (n.d.). *Matplotlib Pie Charts*. [online] Available at: https://www.w3schools.com/python/matplotlib\_pie\_charts.asp.