The main purpose of this project is to show how to clean, process, and manage retail data.

I intend to share knowledge and insight that may be used to create marketing strategies or business plans along this process.

The dataset is a shopping cart database with Australian retail sales data provided from Kaggle, and MySQL is the main analysis tool.

The project is organised as a process of "Question->SQL query->Output->Interpretation" throughout the post.

Dataset link: <https://www.kaggle.com/datasets/ruchi798/shopping-cart-database>

This dataset has four CSV: orders, products, customers, sales

So I made 4 tables for these four CSV documents. (Shown below)

**Customer**

图形用户界面, 文本, 应用程序

描述已自动生成

图形用户界面, 应用程序

描述已自动生成

**Orders**

图片包含 图形用户界面

描述已自动生成

图形用户界面, 文本, 应用程序, 表格

描述已自动生成

**Products**

图形用户界面, 文本, 应用程序

描述已自动生成

图形用户界面

描述已自动生成

**Sales**

图形用户界面, 文本, 应用程序

描述已自动生成

表格

描述已自动生成

This date is from kaggel and the authenticity of this data set is unconfirmed. This dataset was only used as an example to show a demonstration of the skill.

1. **Check the distribution of our customers ( group by state):**

图片包含 图表

描述已自动生成

图形用户界面, 表格

描述已自动生成

South Australia accounts the most in this list and for other states, they account similarly.

**2. Check the distribution of our customers ( group by gender ):**

图形用户界面

中度可信度描述已自动生成

表格

描述已自动生成

Males account for the most in this list, with other genders accounting at a similar rate.

**3. Check the distribution of our customers (group by age):**

图形用户界面, 文本, 应用程序, 电子邮件

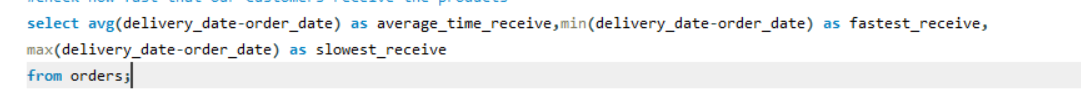
描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

The distribution of the adult age group is generally similar. Age 21-30 and 61-70 are two groups which has relative more people. -

**4. Check how fast our customers receive the products:**



图形用户界面, 文本, 应用程序, Word

描述已自动生成

In general, it needs about 46 days to deliver our products to our customers, however, some customers waited for 99 days.

文本

描述已自动生成

图形用户界面, 应用程序

描述已自动生成

To check who received the order too slowly, we can send some small gifts to apologize.

**5.who is our vip?**

图片包含 文本

描述已自动生成

图形用户界面

描述已自动生成

In this part, we need to analyze who paid the most and if they feel satisfied with the shopping experience.

**6. what product is the popular product?**

图形用户界面

低可信度描述已自动生成

表格

中度可信度描述已自动生成

Also analyze the popular product (depending on the payment)

Denim with indigo is a very famous product, however, the delivery time is too long.

**7. Which products were sold the most in the last month (October)?**

图形用户界面, 文本, 应用程序

描述已自动生成

表格

描述已自动生成

In October, the orange trench coat in size M was sold the most, which is followed by the flannel shirt and the puffer jacket in the small size.

**8. How did the revenue change in the last 4 quarters?**

表格

描述已自动生成

表格

描述已自动生成

,

Because our data didn’t have enough records in November and December, the revenue in the last quarter is the least. So, the last quarter doesn’t have too much

reference value

**Conclusion:**

Our clients have distributed almost equality in all states, all genders, and all adult age groups. Age 21-30 and 61-70 have relatively more customers so we should customize more for these two age groups. But obviously, we need to provide better delivery service, because even our VIP customers need to wait for around 60 days to receive the products. Throughout the whole year, Denim with indigo, blue, violet, and orange are popular products; the first quarter and last quarter also got the most revenue so we can have more discounts in these two periods of time. In the last month (October), trench coats, flannel, and puffers are more popular.