Descriptive analysis on E-cart company data

Business Task

E-cart business head wants our data analytics team to perform descriptive analysis in order to find out how the business has been running. As a final deliverable, we are expected to provide the insights in the form of an interactive and dynamic dashboard that displays the below information and KPI's.

The interactive & dynamic dashboard should include :

- Average delivery time required by our company to deliver products in each state.
- The most preferred ship mode and delivery time taken by them.
- Revenue generated from each state.
- Revenue for different quarters and years.
- Quantity bought and Revenue generated from each customer
- Quantities sold and Revenue generated from different subcategories and categories separately.

DATA COLLECTION

Data has been collected from a Udemy course which is made available provided by **Start-Tech Academy**.

Data set link : Click Here

There are 3 files namely : SALES, CUSTOMERS AND PRODUCTS.

DATA WRANGLING

Data was cleaned using PSQL and Excel. I cleaned data for further analysis in the following ways :

Functions like **+month(date)** and **+year(date)** were used to extract month and year respectively in new columns.

- New column called Delivery time taken was calculated using SQL epoch function as shown below :

```
select *, round((extract(epoch from ship_date)-
extract(epoch from order_date))/86400) as
delivery_time
from sales
```

Values from some columns were formatted to numeric format using Excel menu options.

Data Validation Checks performed using excel menu:

- Checked for any null values :
- Checked for any duplicates.
- checked for spelling errors.

DATA ANALYSIS

Descriptive analysis was done using PSQL to fetch insights from data and answer the business requirements.

Creating Tables

Creating Sales Table and importing corresponding data

```
create table Sales (
Order_Line int,
Order_ID varchar,
Order_Date date,
Ship_Date date,
Ship_Mode varchar,
Customer_ID varchar,
Product_ID varchar,
Revenue float,
Quantity int,
Delivery_time int
);
copy Sales from 'C:\Program
Files\PostgreSQL\15\Sales.csv' delimiter ',' header;
```

Creating Product Table and importing corresponding data

```
create table product (
product_id varchar,
category varchar,
sub_category varchar,
product_name varchar);
copy product from 'C:\Program
Files\PostgreSQL\15\product.csv' delimiter ',' header;
```

Creating Customer Table and importing corresponding data

```
create table customer (
customer_id varchar,
customer_name varchar,
segment varchar,
age int,
country varchar,
city varchar,
state varchar,
postal_code int,
region varchar
);
copy customer from 'C:\Program
Files\PostgreSQL\15\Customer.csv' delimiter ','
header;
```

Descriptive analysis

Average delivery time required by our company to deliver products in each state.

```
select state, round(avg(delivery_time))
from customer
join sales on customer.customer_ID=sales.customer_id
group by state;
```

The most preferred ship mode and delivery time taken by them.

```
select ship_mode, round(avg(delivery_time))
from sales
group by ship_mode
```

- Revenue generated from each state.

```
select sum(revenue) as "revenue by state", state
from customer
join sales on customer.customer_id=sales.customer_id
group by state
```

- Quantity bought and Revenue generated from each customer

```
select sum(quantity) as "total items bought",
sum(revenue) as revenue_generated, customer
from customer
join sales on customer.customer_id=sales.customer_id
group by customer
```

- Revenue for different quarters and years.

```
select year, sum(revenue) as revenue_generated
from sales
group by year
```

Quantities sold and Revenue generated from different subcategories

```
select sub_category, sum(Quantity) as
Quantity_of_subcategory_products_sold, sum(revenue) as
revenue_from_sub_categories
from product
join product on product.product_id=sales.product_id
group by sub_category
```

Quantities sold and Revenue generated from different Categories

```
select category, sum(Quantity) as
Quantity_of_category_products_sold, sum(revenue) as
revenue_from_categories
from product
join product on product.product_id=sales.product_id
group by category
```

VISUALISATION

Reporting

It can be concluded that on an average the time required to deliver products in almost all state was around 4-6 days.

The majority of products were shipped by standard mode which takes around 6 days for delivery of products and comparatively less were shipped by other modes, especially same day delivery was ordered by least customers. It can be concluded that generally people aren't in rush to get their products as soon as possible unless they are in urgent need of that product. People are okay with standard delivery as long as it cost's less or is free.

We have generated huge revenue from states who seem to have comparatively high population, namely New York, California and Texas. Some of the states from which we generated the least revenue are Kansas, columbia etc. It can be concluded that the states with top economies and high population are generating the most revenue for us. So, its best to prioritize our business in this states.

The above chart clearly shows that it's not necessarily true that if the quantities bought by a customer is more than the revenue generated from them will also be more. Revenue is also dependent on which product a person buys rather than just the quantity of any product. We can filter by the customer names to find out the quantities bought and revenue generated from them. This way we can find our top 5 customers.

There's a trend which we can see every year and that is the revenue is slowly growing from 1st quarter to 2nd quarter and then from 2nd quarter to 4th quarter revenue growth was steep. Except in year 2016, the steep increase of revenue occurred from 3rd quarter. It can also be observed that, revenue generated in 2016 & 2017 was more than previous years. Looking at the quarterly revenue, we can infer that people buy our products a lot more in rainy & winter season, probably because there's 2 month holiday in summer season and thus, only few people buy office supplies in that time period.

The sub-categories which not only were sold the most but also the generated the most revenue are chairs from furniture category, phones from technology category and storage from office supplies category. So, it's important to ensure that we have stock of this sub-category products. On the other hand, some sub-category products from office supplies category generated the least revenue as they are inexpensive and generally less in demand. To name a few such office supplies include fasteners, labels and envelops.

Technology products are almost always expensive. Thus, from the above graph we can see technology products generated the most revenue, even though they were sold the least. We can see from the graph how furnitures and office supplies perform as well.

SALES DASHBOARD

This is an interactive dashboard which stakeholders can use to filter by categories, states, customer name and year to drill down to the exact data they want to find. The **KPI sales dashboard** is also dynamic, so it will automatically update itself if the data source is updated.

Stakeholders can easily interact with this dashboard to understand how business is performing and gain insights from it.