

[Part 2] Task 01. Create a new notebook, import the necessary analysis and visualization libraries, then import your most up-to-date project data

Import Libraries

```
In [243]: import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
import seaborn as sns
import scipy

In [249]: # Set path
path = r'C:\Users\Aasus\Music\Instacart Basket Analysis'

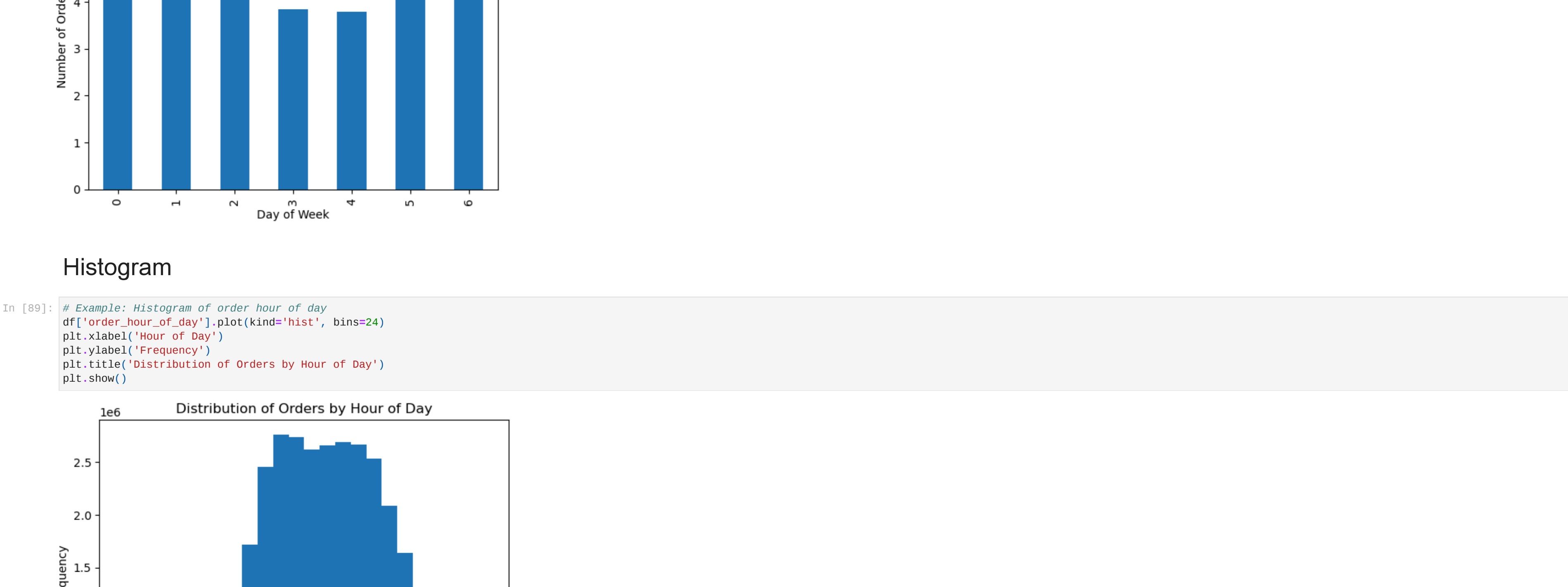
In [231]: # Import most up-to-date project data
instacart = pd.read_csv(os.path.join(path, 'Data', 'Prepared Data', 'orders_cleaned.csv'))

In [58]: # Import most up-to-date project data
# Define the path to the dataset
file_path = r'C:\Users\Aasus\Music\Instacart Basket Analysis\Data\Prepared Data\ords_prods_merge_final.pkl'

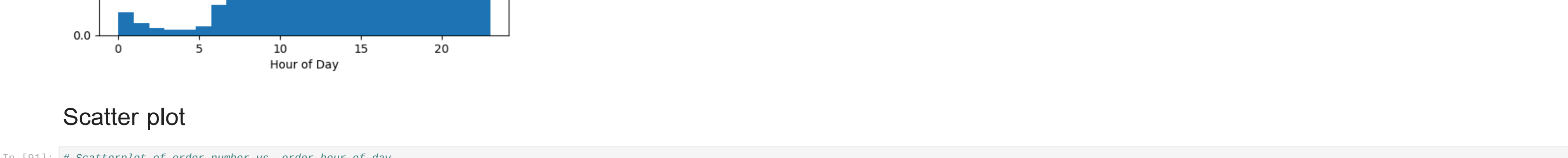
# Load the dataset
df = pd.read_pickle(file_path)
```

[Part2] Task 02. If you haven't done so already, follow the instructions in the Exercise to create a bar chart, a histogram, a scatterplot, and a line chart, using a sample subset for the line chart.

Bar Chart



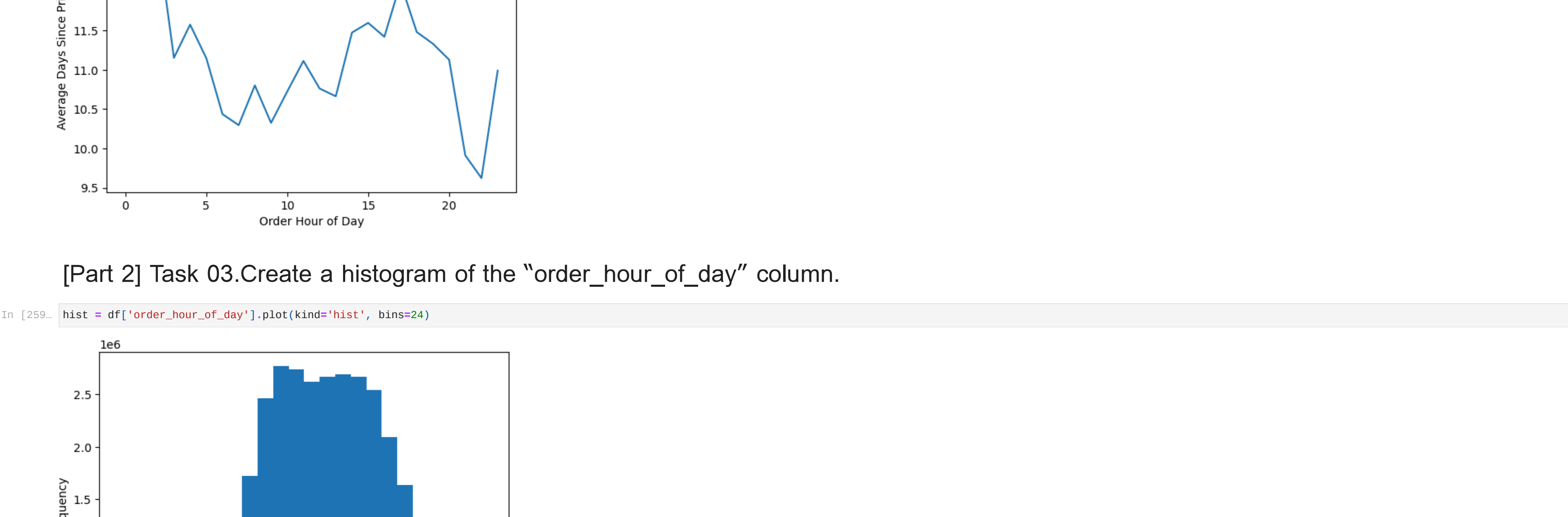
Histogram



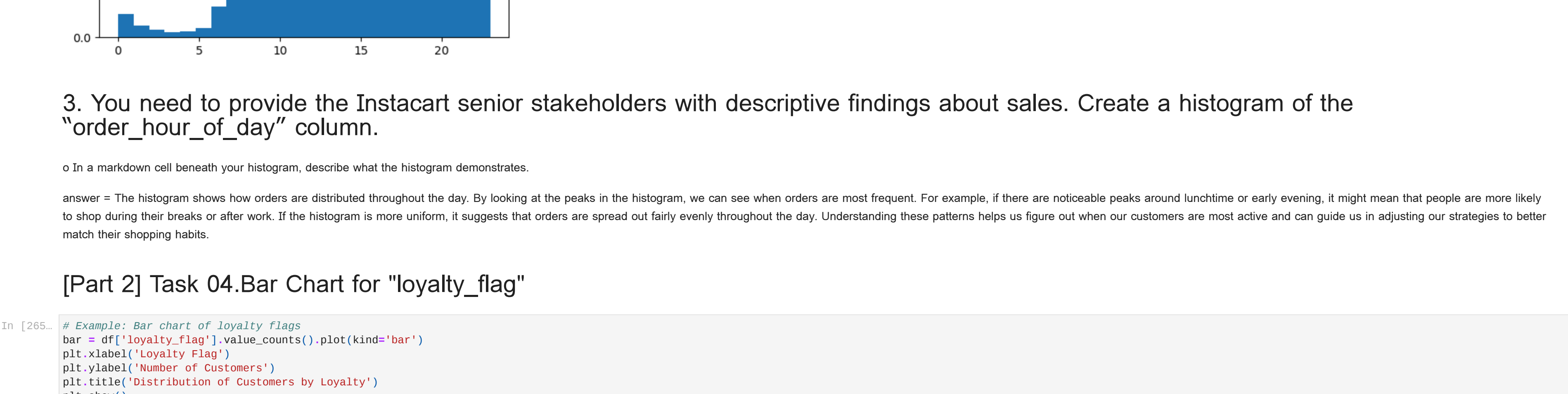
Scatter plot



Line Chart



[Part 2] Task 03.Create a histogram of the “order_hour_of_day” column.

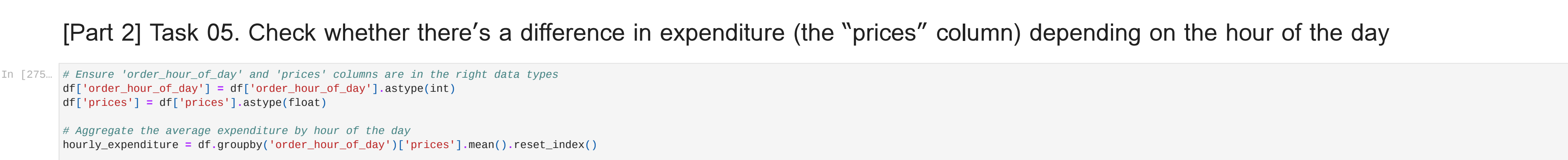


3. You need to provide the Instacart senior stakeholders with descriptive findings about sales. Create a histogram of the “order_hour_of_day” column.

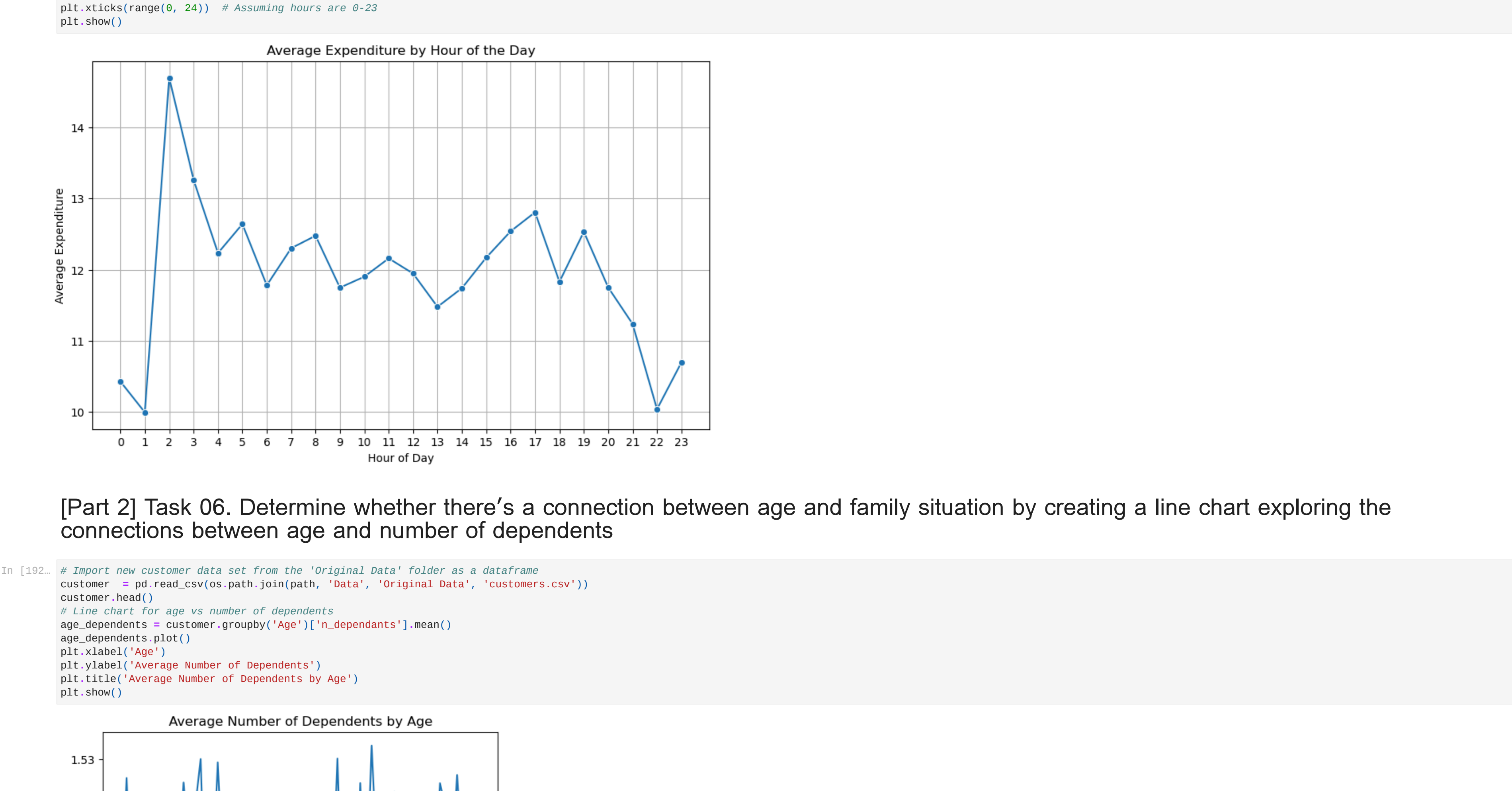
o In a markdown cell beneath your histogram, describe what the histogram demonstrates.

answer = The histogram shows how orders are distributed throughout the day. By looking at the peaks in the histogram, we can see when orders are most frequent. For example, if there are noticeable peaks around lunchtime or early evening, it might mean that people are more likely to shop during their breaks or after work. If the histogram is more uniform, it suggests that orders are spread out fairly evenly throughout the day. Understanding these patterns helps us figure out when our customers are most active and can guide us in adjusting our strategies to better match their shopping habits.

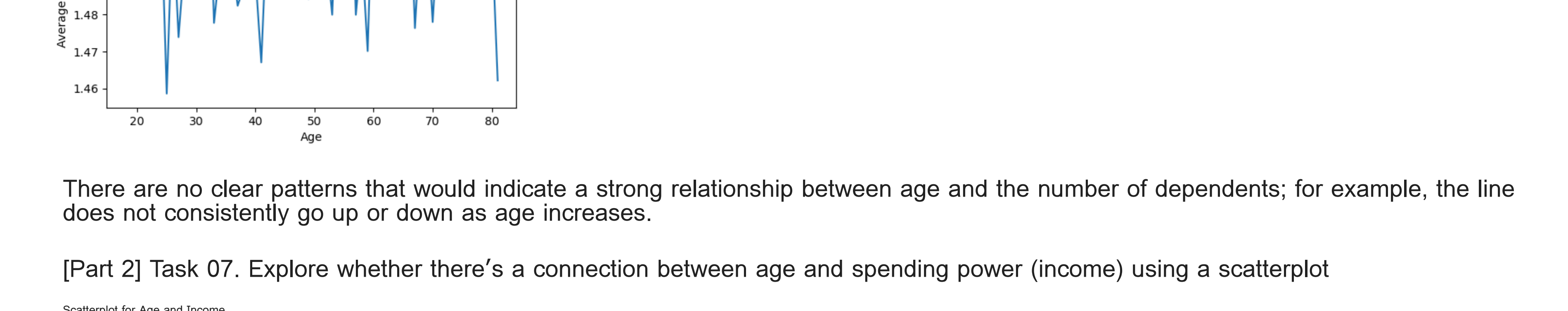
[Part 2] Task 04.Bar Chart for "loyalty_flag"



[Part 2] Task 05. Check whether there’s a difference in expenditure (the “prices” column) depending on the hour of the day

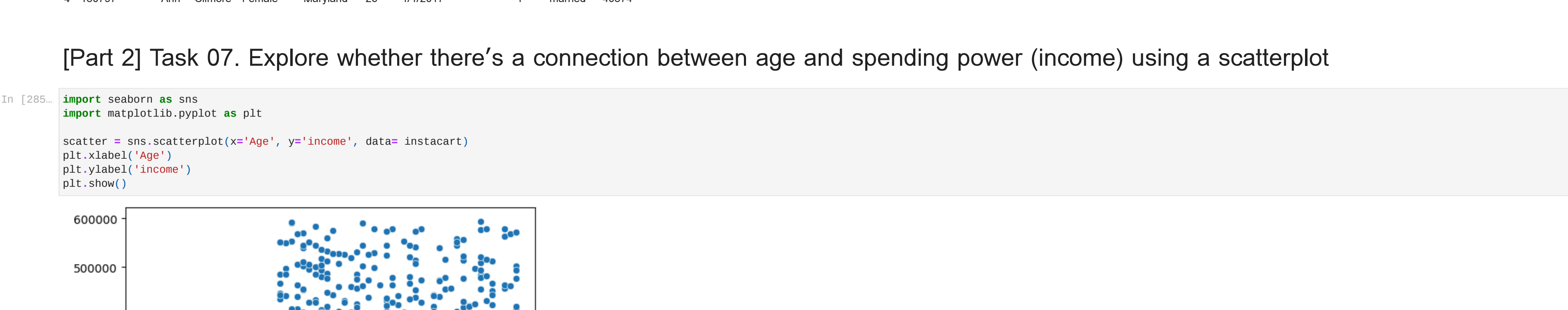


[Part 2] Task 06. Determine whether there’s a connection between age and family situation by creating a line chart exploring the connections between age and number of dependents

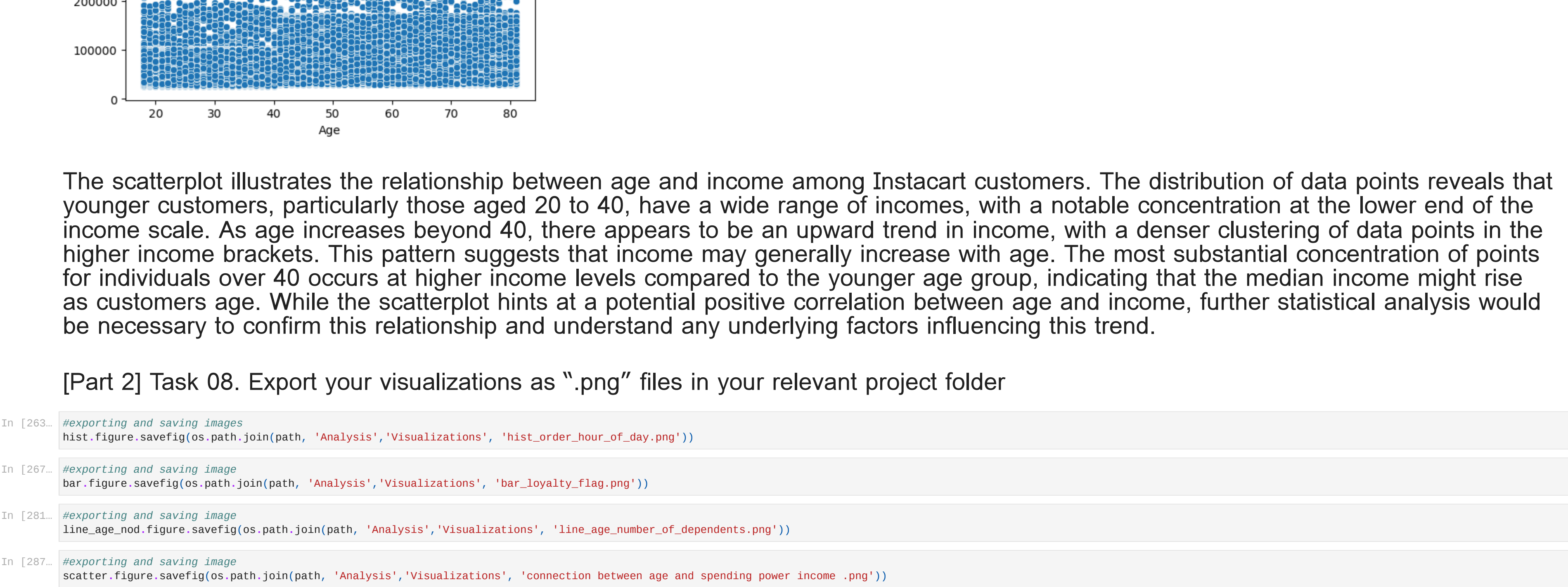


There are no clear patterns that would indicate a strong relationship between age and the number of dependents; for example, the line does not consistently go up or down as age increases.

[Part 2] Task 07. Explore whether there’s a connection between age and spending power (income) using a scatterplot



[Part 2] Task 07. Explore whether there’s a connection between age and spending power (income) using a scatterplot



The scatterplot illustrates the relationship between age and income among Instacart customers. The distribution of data points reveals that younger customers, particularly those aged 20 to 40, have a wide range of incomes, with a notable concentration at the lower end of the income scale. As age increases beyond 40, there appears to be an upward trend in income, with a denser clustering of data points in the higher income brackets. This pattern suggests that income may generally increase with age. The most substantial concentration of points for individuals over 40 occurs at higher income levels compared to the younger age group, indicating that the median income might rise as customers age. While the scatterplot hints at a potential positive correlation between age and income, further statistical analysis would be necessary to confirm this relationship and understand any underlying factors influencing this trend.

[Part 2] Task 08. Export your visualizations as “.png” files in your relevant project folder

```
In [283]: #Exporting and saving images
hist.figure.savefig(os.path.join(path, 'Analysis','Visualizations', 'hist_order_hour_of_day.png'))

In [287]: #Exporting and saving image
bar.figure.savefig(os.path.join(path, 'Analysis','Visualizations', 'bar_loyalty_flag.png'))

In [281]: #Exporting and saving image
line_age_mod.figure.savefig(os.path.join(path, 'Analysis','Visualizations', 'line_age_number_of_dependents.png'))

In [287]: #Exporting and saving image
scatter.figure.savefig(os.path.join(path, 'Analysis','Visualizations', 'connection between age and spending power income .png'))

In [ ]:
```