

PowerSportsNation Analysis

By: Morgan and Randy

2/9/2023

Overview

Topic and Reasoning

For this project, we are working with PowerSportsNation, a power sports dealer located in Nebraska. PowerSportsNation provides quality used parts that are cheaper than retail. They have asked us to analyze their data for trends and provide possible business recommendations.

We chose this dataset to work with as Randy currently works with the company and provides a great opportunity to work with an actual company and provide real world solutions to improve their company.

Data Description

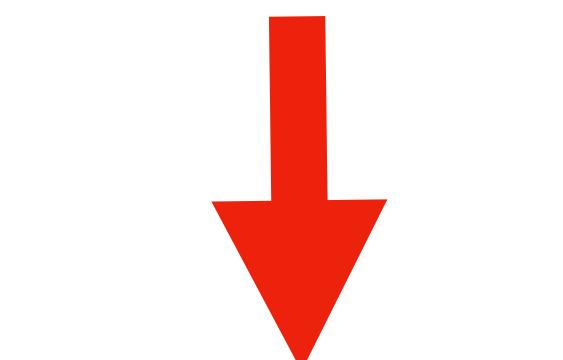
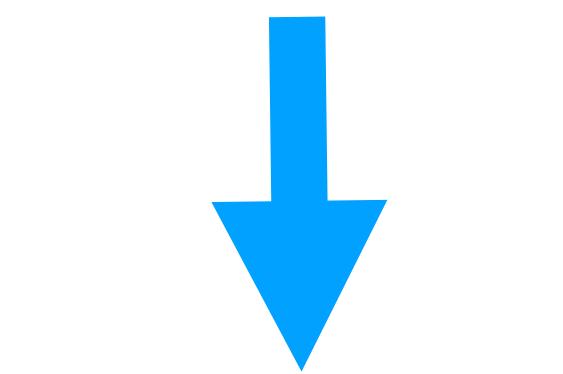
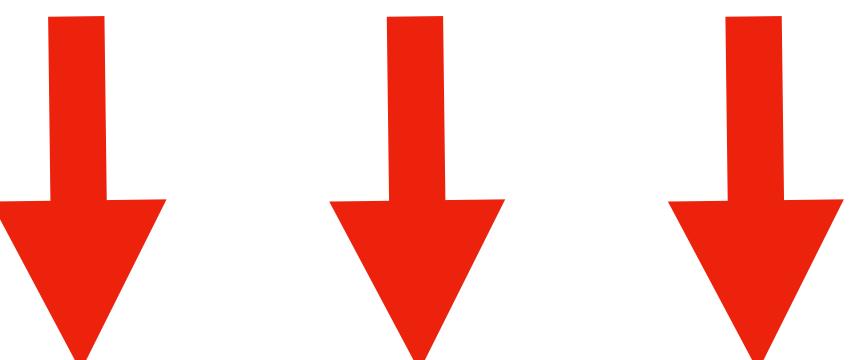
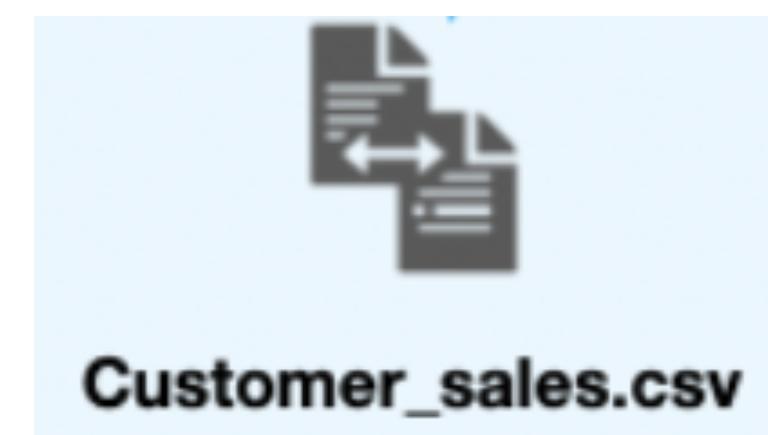
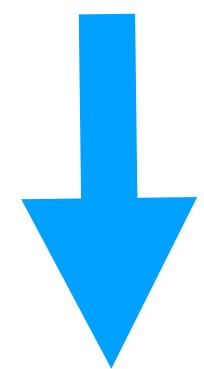
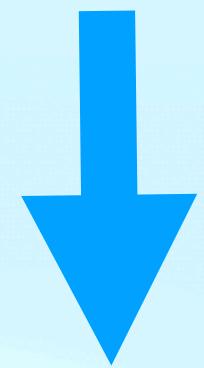
- PSN works with used parts that are sold across the world. To understand customer segmentation we took two years of customer order details per customer, such as customer ID, part brand, the market, and line of business.
- In addition, to better understand customer segmentation, we needed to collect zip code demographics, such as zip code, latitude, and longitude.
- Then we can merge all the data together to get a more comprehensive analysis of customer segmentation based on customer behavior and location demographics.

What Do We Want to Know?

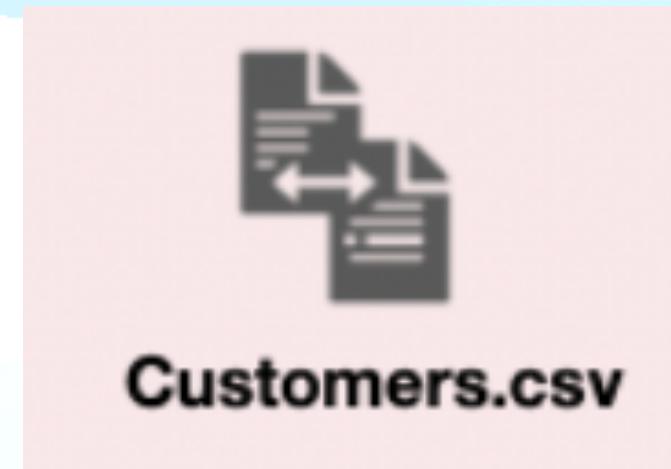
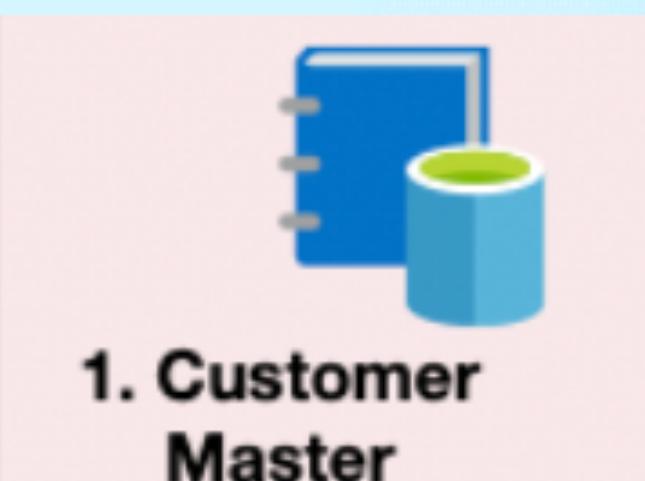
Questions we plan to answer

- What are the demographics and behaviors of customer segmentation? Can we develop marketing personas?
 - Who are our core dealers/customers that the company should be reaching out to for business?
 - How much do one time purchasers provide to the company?

EXTRACT



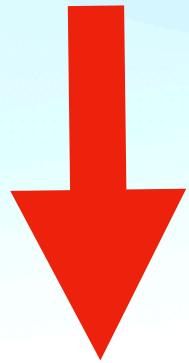
TRANSFORM



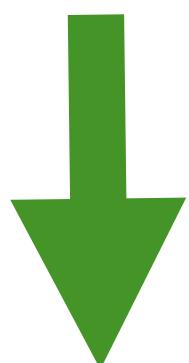
TRANSFORM



4. Customer_Data_Merge



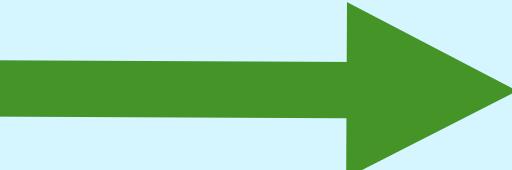
Customer_date_merged.csv



LOAD

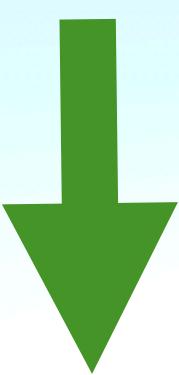


Azure DB



Machine Learning

Charts &
Unstructured
Learning
Classifications



Dashboard and
Charts

Exploration of the Data

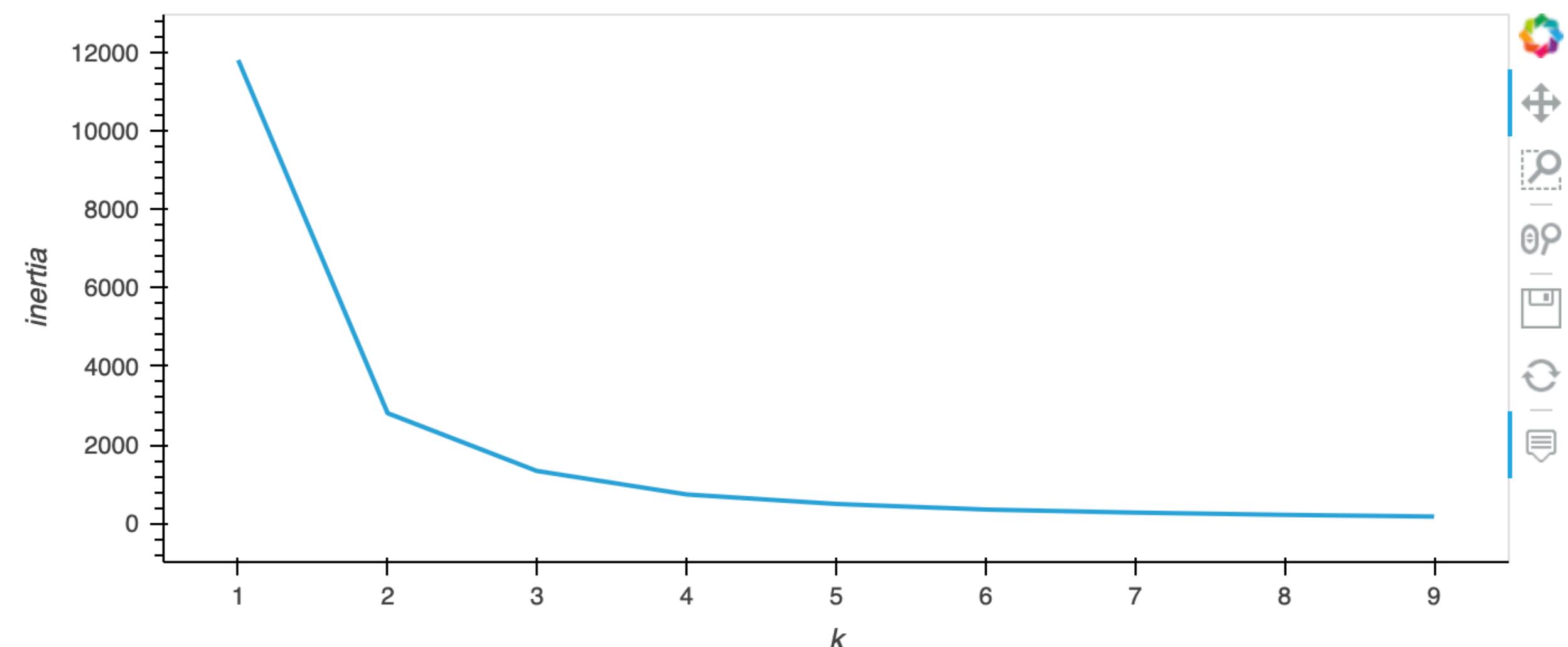
Items Used: Jupyter Notebook, Python, Pandas, KMeans, Seaborn

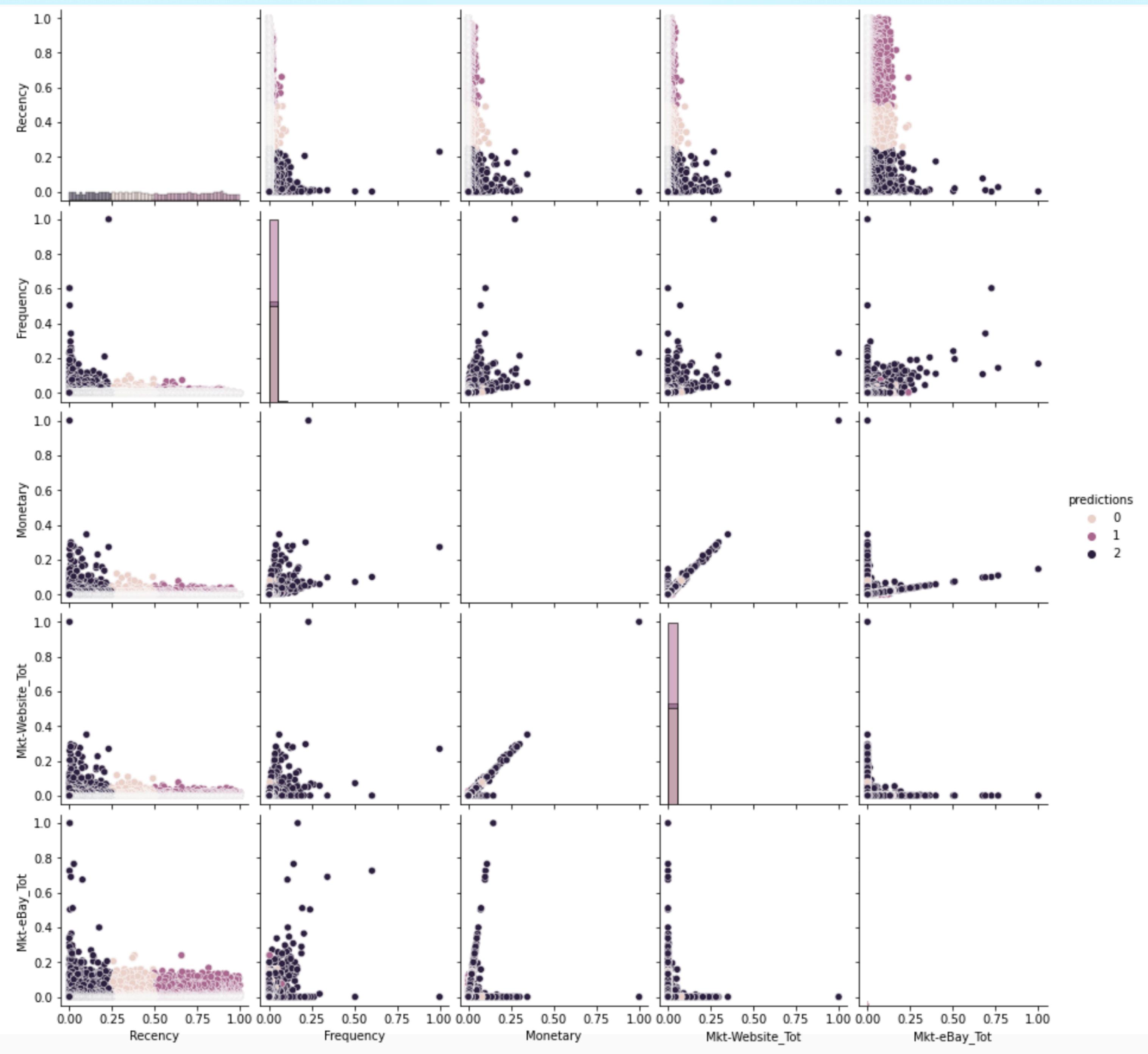
- To gather a basic understanding of the data, we created a DataFrame with the following columns:

- Recency
- Frequency
- Monetary
- Market Website Total
- Market eBay Total

```
sse = {}
K = range(1, 10)
for k in K:
    kmeanmodel = KMeans(n_clusters=k).fit(data_df_scaled)
    sse[str(k)] = kmeanmodel.inertia_
# print(sse)

# Plot
elbow_curve_df = pd.DataFrame({"k":sse.keys(), "inertia":sse.values()})
elbow_curve_df.hvplot.line(x="k", y="inertia")
```





Analysis

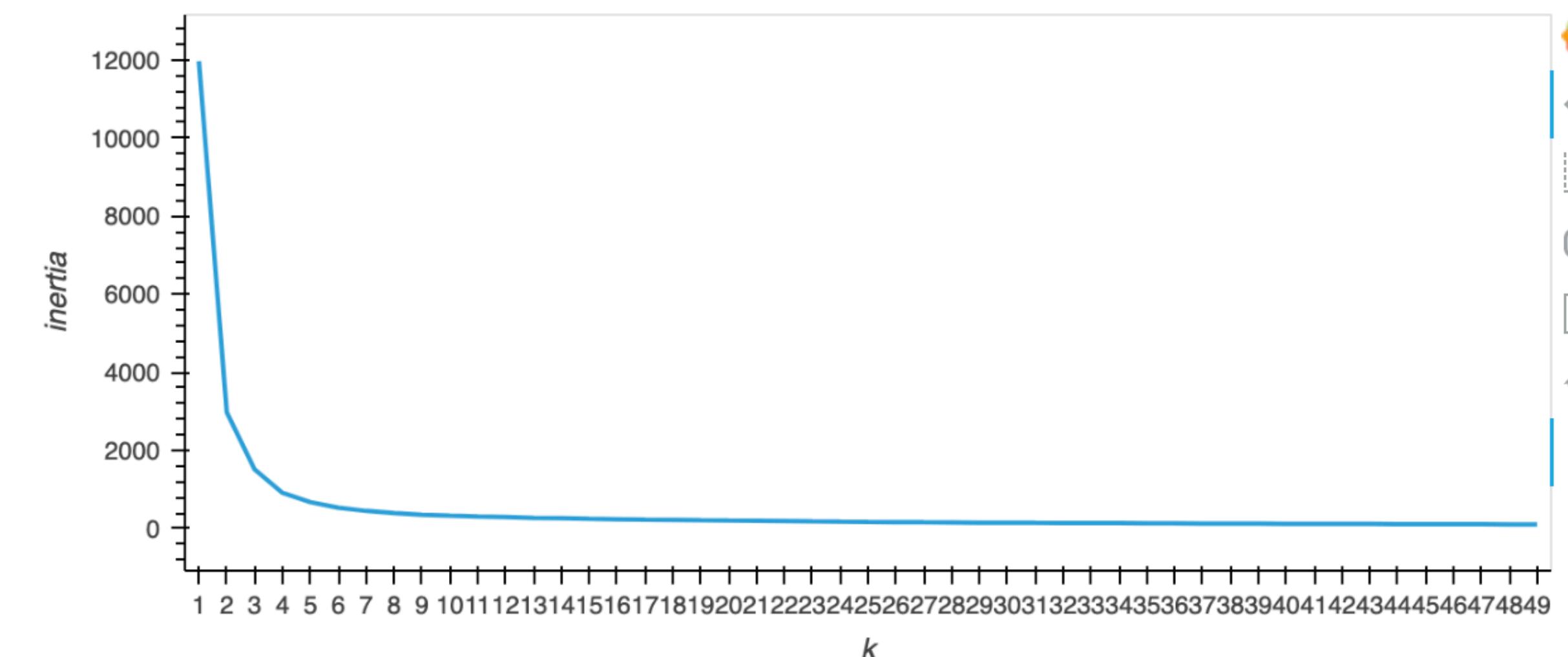
Items Used: Jupyter Notebook, Python, Pandas, KMeans, Seaborn

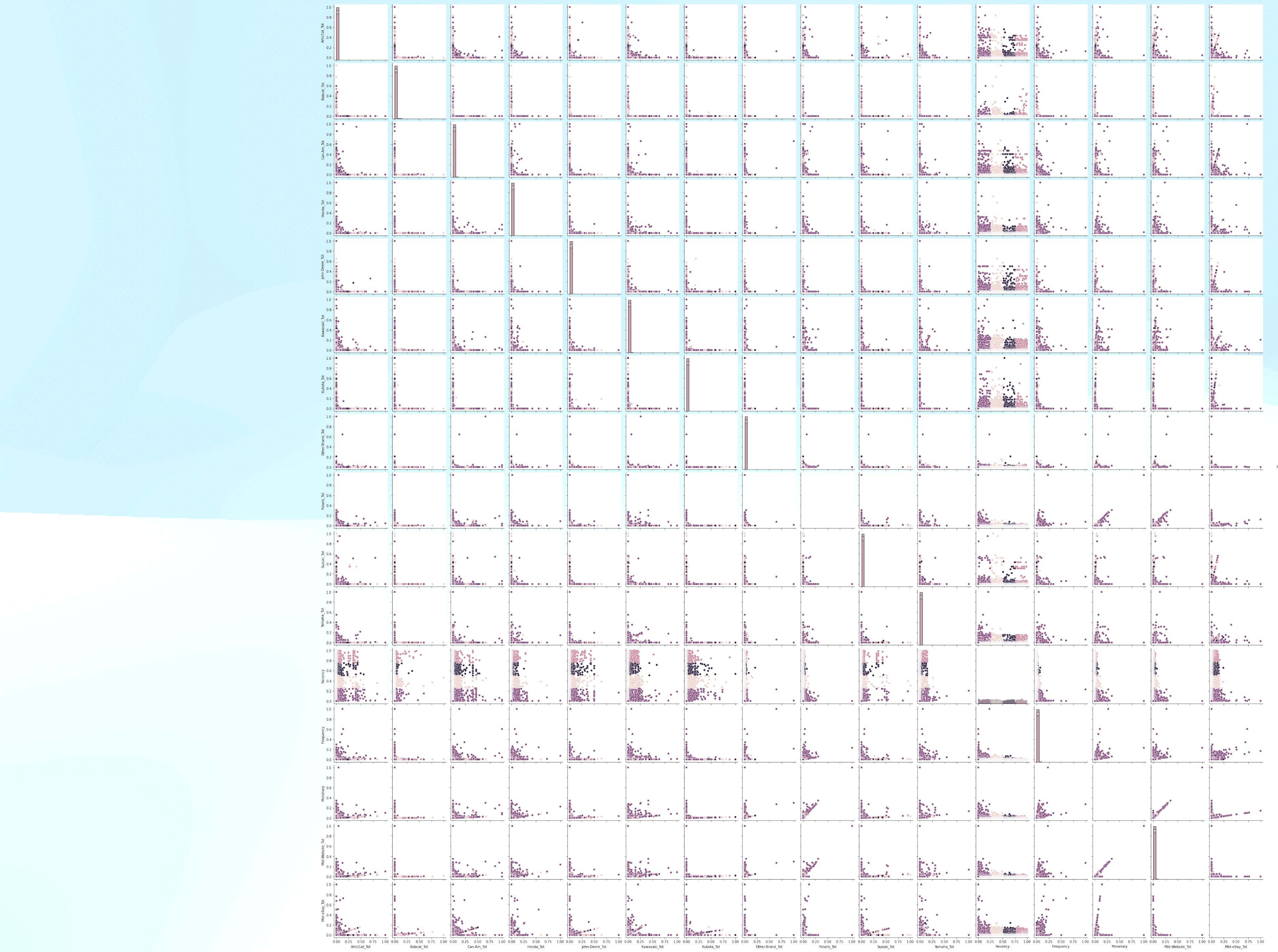
- Using the same columns as previously, for the analysis we included the brands for the analysis.

- For this analysis, we also increased the range to get a better cluster value

```
sse = {}
K = range(1, 50)
for k in K:
    kmeanmodel = KMeans(n_clusters=k).fit(data_df_scaled)
    sse[str(k)] = kmeanmodel.inertia_
# print(sse)

# Plot
elbow_curve_df = pd.DataFrame({"k":sse.keys(), "inertia":sse.values()})
elbow_curve_df.hvplot.line(x="k", y="inertia")
```





Challenges

- Limitation to the data
- Merging the datasets together

Limitations

- Data is dependent on the company, employees, and external sellers such as Amazon and ebay