Case Study

Instacart Grocery Basket Analysis

Problem: Sales are good but the VP wants to make sales better.

Goal: Uncover information about sales patterns by performing exploratory analysis to gain insights on customer demographics and buying trends in order to target customers with applicable marketing strategies.

Role: Data Analyst

Stakeholders: Vice President of Marketing, Sr. Vice President of Sales, Instacart Customers.

Key Questions:

- What are the busiest/least busy days of the week and hours of the day in order to schedule ad placement?
- Is there a particular time of day when people spend the most money, as this might inform the type of products they advertise during these times?
- Instacart has a lot of products with different price tags. Marketing and sales want to use simpler price range groupings to help direct their efforts.
- Are there certain types that are more popular than others?
-

Tools Used:

Data

Data was provided by CareerFoundry and had the following datasets:

1. Orders
2. Products
3. Customers
4. Departments

More information on datasets can be found here.

Python Skills

- Importing libraries and datasets
- Data Cleaning
- Descriptive Analysis
- Data Wrangling
- Grouping and Aggregating data
  - GroupBy and For Loops
  - CrossTabs
  - Data Aggregation
- Visualizations with Python libraries
  - Histogram
  - Bar Chart
  - Line Graph
- Exporting datasets

Objective:

Perform exploratory to answer business questions and derive insights about buying trends and customer demographics in order to target customers with applicable marketing strategies.

Data Cleaning & Wrangling

Import Datasets into Python using Jupyter Notebook.

- > 32 million records
- 4 different datasets

Data Cleaning & Deriving New Variables

Clean and wrangle data

Gain understanding of data through:

- Renaming columns
- Removing columns
- Changing data types to decrease RAM
- Clean up inaccurate data

Filter and Aggregate Data

Obtain descriptive statistics

Derive Variables & Group Data

Create new variables and flags to aid in profiling customers.

Filter and Aggregate Data by:

- Groupby
- Aggregate function
- CrossTab function

Data Exploration & Analysis

Create visualizations in Python using:

- Matplotlib
- Seaborn
- Scipy

Presenting Results

Utilize Excel to present Final Report

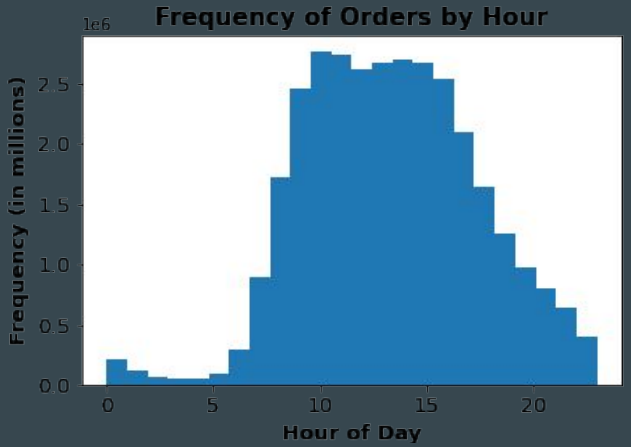
Project posted on Github Repository

Tools Used:

Initial Exploration: Order Frequency

Busiest Days: Saturday and Sunday
Least Busy: Tuesdays/Wednesdays

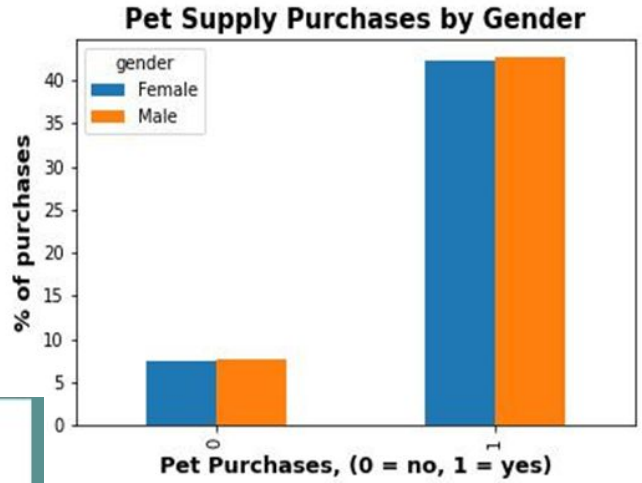
Hour of the Day: Least busy from Midnight - 7am.



Creating Customer Profile to Market Pet Supplies

Step 1: Create a pet flag
0 = no pet supplies purchased
1 = pet supplies purchased

Step 2: Aggregate the Data



7c. Pet flag

```
# first, identify all the products purchased from the pet department
df3['pet_stuff'] = [1 if x == 'pets' else 0 for x in df3['department']]
```

```
# then, apply it across the whole customer's orders
df3['pet_stuff'] = df3.groupby(['customer_id'])['pet_stuff'].transform('max')
```

```
# reduce datatype to save RAM
df3['pet_stuff'] = df3['pet_stuff'].astype('int8')
```

Digging Deeper:

Which states have the highest purchases of pet supplies?

Pennsylvania, California, and Georgia!

- In order to do this, I created a flag for customers who purchased pet products and labeled the flag as a “pet owner.”
- Then I pulled state and pet owner values to see what states had the highest pet owner values.

```
state          pet_owner
Pennsylvania   yes          165620
California     yes          161600
Georgia        yes          161059
Arizona        yes          160533
Rhode Island   yes          160454
dtype: int64
```

```
# creating flag for just customers that have a 1 for purchasing pet supplies
df.loc[df['pet_stuff'] == 1, 'pet_owner'] = 'yes'
df['pet_owner'].value_counts(dropna=False)
```

```
yes    7882948
NaN    1387072
Name: pet_owner, dtype: int64
```

```
# top 5 states that purchase pet products
top5_state_pet = df[['state', 'pet_owner']].value_counts().sort_values(ascending=False).head()
top5_state_pet
```

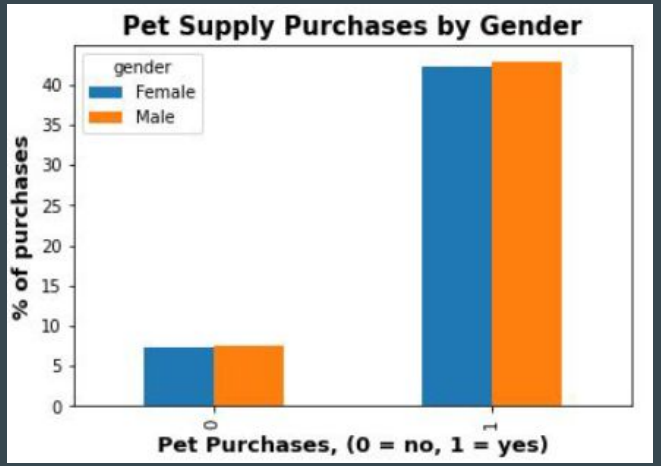
.....continued digging:

Is there a difference in gender when purchasing pet supplies?

There appears to be no difference in gender amongst pet supply purchases.

```
# percent of males/females who purchase pet products
crosstb_gender_pets = round(pd.crosstab(df['pet_stuff'],df['gender'], normalize=True)*100,2)
```

gender	Female	Male
pet_stuff		
0	7.37	7.59
1	42.29	42.74



Final Report

Recommendations:

- Schedule ads during the slowest days of the week which are Tuesday and Wednesday from Midnight - 7:00 am. This is when fewest orders are placed.

- There is no gender difference when it comes to pet supply purchases; however the top 5 states to focus marketing attempts for pet supplies would be:
  - Pennsylvania
  - California
  - Georgia
  - Arizona
  - Rhode Island

Deliverables

- Github Repository Link
  - Project Brief
  - Code/Scripts
  - Final Report for much more information!