EXCERPT

INSTACART

A successful online grocery store that operates through an app. Executives are curious to learn about different customer segments and understand their purchasing behavioral patterns for applicable marketing campaigns.

NOTE: Instacart is a real company that's made its data publicly available online. However, the content of this project has been fabricated for the purposes of this Achievement.

OBJECTIVE

Produce an exploratory analysis of consumer behavior and sales patterns in sustaining information that will benefit sales and marketing departments in developing tactical promotional efforts.

PROJECT & DATA

- Project Brief
- Customer Data Set, Customer Orders, Products, and Department | *Modified Open Source from Instacart*
- Data Dictionary | *Link provided by CareerFoundry*

LIMITATIONS

- Data only contains records from 2017.
- Customer demographics are limited to income, age, family size, and marital status.

TECHNIQUES APPLIED

- Data Cleaning: Wrangling and Subsetting
- Data Consistency Checks
- Combining and Exporting Data
- Deriving New Variables
- Grouping Data and Aggregating Variables
- Python Visualization and Excel Report

TOOLS

DATA METHODOLOGY

ORGANIZING DATA

Tracking data irregularities like mixed-type variables, missing and/or duplicate values for ratification. Methodically noting steps trailing the transformation of the dataframe will allow other users to identify procedural actions taken.

DERIVING VARIABLES

Using conditional logic in the form of if-statements, user-defined functions, the loc() function, and for-loops in deriving new columns.

AGGREGATING DATA

Creating flags and placing new columns for a summary of descriptive analysis with groupby() function.

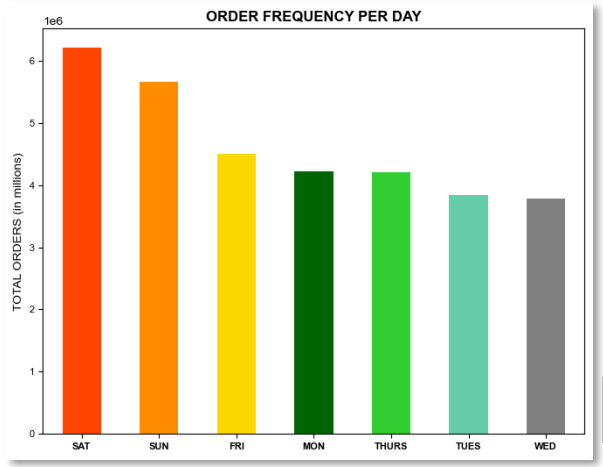
PYTHON VISUALIZATION & EXCEL REPORT

A merged dataframe is utilized to generate a compelling visualization addressing the stakeholders' questions.

BUSINESS ANALYSIS

ANALYTICAL & VISUALIZATION TOOLS APPLICATION

Saturday, **Sunday**, and **Friday** are the **busiest** days for shoppers.

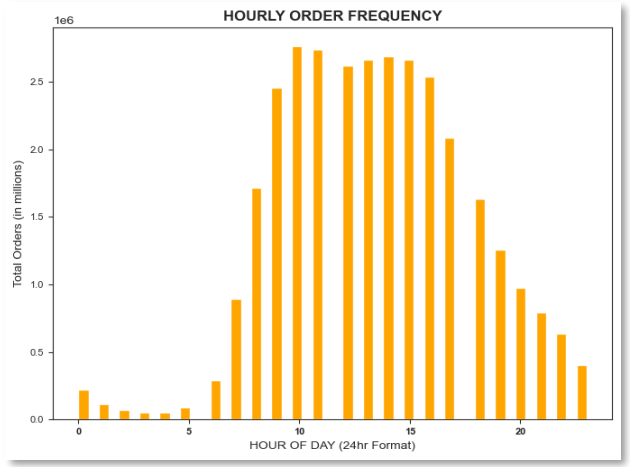



```
bar_chart = final_df['order_day_of_week'].value_counts().plot.bar(color = ['orangered', 'di

# Adding Chart Labels
sns.set_style('ticks')
bar_chart.set_xticklabels(['SAT', 'SUN', 'FRI', 'MON', 'THURS', 'TUES', 'WED'])
plt.ylabel('TOTAL ORDERS (in millions)')
plt.xticks(fontsize = 8, fontweight = 'bold')
plt.yticks(fontsize = 8)
plt.title('ORDER FREQUENCY PER DAY', fontweight = 'bold')
plt.show()
```

FIG. 4a

Purchasing frequency **increases** between 9 am through 3 pm.




```
hist2 = final_df['order_hour_of_day'].plot.hist(bins = 50, color = 'orange', figsize = (8,6))

# Adding Chart Labels
sns.set_style('ticks')
plt.ylabel('Total Orders (in millions)')
plt.xlabel('HOUR OF DAY (24hr Format)')
plt.xticks(fontsize = 8, fontweight = 'bold')
plt.yticks(fontsize = 8)
plt.title('HOURLY ORDER FREQUENCY', fontweight = 'bold')
plt.show()
```

FIG. 4b

There are **21 departments** that offer **multiple price** ranges.

department	HIGH	LOW	MID
alcohol	-	33,046	111,581
babies	-	121,484	288,908
bakery	-	274,986	845,842
beverages	-	814,697	1,757,204
breakfast	-	209,185	461,665
bulk	-	1,181	32,270
canned goods	-	281,711	730,363
dairy eggs	4,877	1,370,908	3,801,397
deli	-	299,220	704,614
dry goods pasta	-	284,346	537,790
frozen	-	647,617	1,474,114
household	-	245,070	454,787
international	-	73,203	182,788
meat seafood	392,855	-	281,926
missing	-	18,516	46,252
other	-	15,259	19,152
pantry	221	504,521	1,277,963
personal care	-	123,874	300,432
pets	-	28,165	64,895
produce	-	2,585,708	6,493,565
snacks	-	1,742,143	1,024,263

```
#16 Department Price Range and Check Output

dept_prc_rnge = pd.crosstab(df_cstmr_prfl['department'], df_cstmr_prfl['price_range'])

dept_prc_rnge
```

FIG. 4c

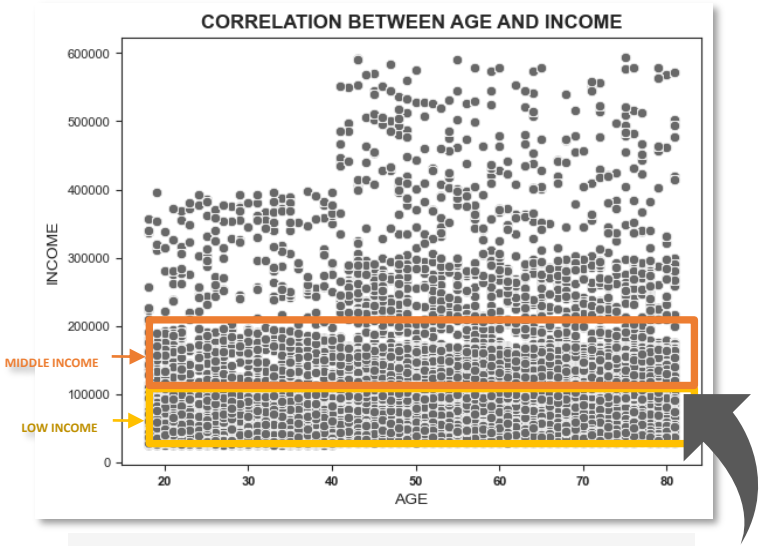
VIEW FULL REPORTS BELOW

Python Codes | Final Excel Report

CONSUMER SEGMENT ANALYSIS

ANALYTICAL & VISUALIZATION TOOLS APPLICATION

The customers' **income** ranges between **low to mid-range** describing probable spending power.



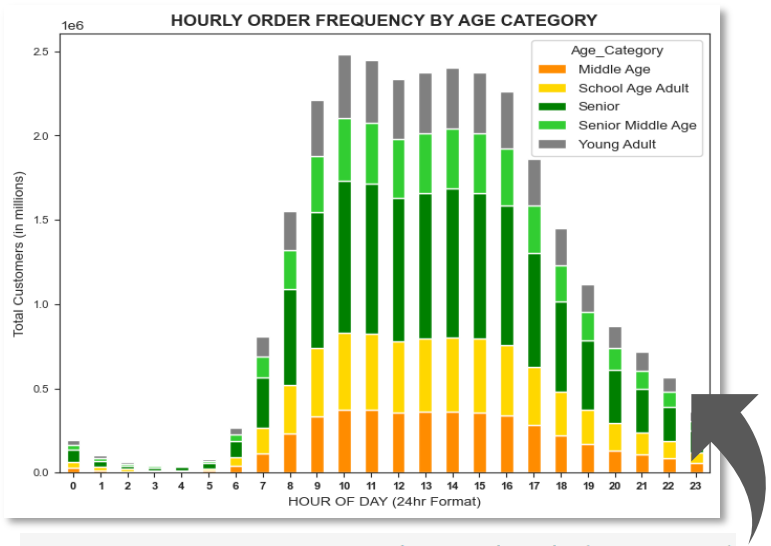

```
# Creating Scatterplot2 for spending power by age group

scatterplot2 = sns.scatterplot(x = 'Age', y = 'Income', data = final_df)

# Adding Chart Labels
sns.set_style('ticks')
plt.ylabel('INCOME')
plt.xlabel('AGE')
plt.xticks(fontsize = 8, fontweight = 'bold')
plt.yticks(fontsize = 8)
plt.title('CORRELATION BETWEEN AGE AND INCOME', fontweight = 'bold')
plt.show()
```

FIG. 4d

Seniors appear to be **frequent shoppers**.



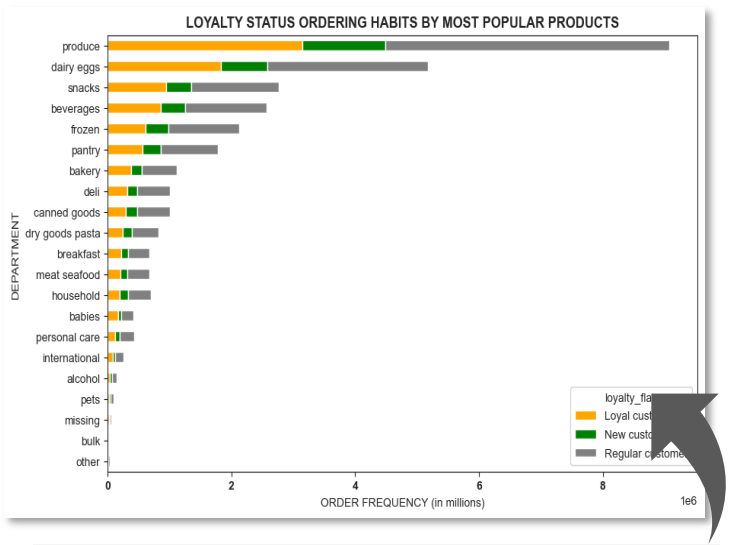

```
#2a Create a Stacked Bar for Age Category Profile of Shopping habits (Order Hour of Day)

stckbr_age_shppngtm_hbts = Age_shppngtm.plot(kind = 'bar', stacked = True, color = ['darl

# Adding Chart Labels
sns.set_style('ticks')
plt.ylabel('Total Customers (in millions)')
plt.xlabel('HOUR OF DAY (24hr Format)')
plt.xticks(fontsize = 8, fontweight = 'bold')
plt.yticks(fontsize = 8)
plt.title('HOURLY ORDER FREQUENCY BY AGE CATEGORY', fontweight = 'bold')
plt.show()
```

FIG. 4e

Produce, **dairy eggs**, and **snacks** are the most **profitable** goods.




```
#11a Create a Bar Chart for department distribution by Loyalty Status

br_dept_lylty = dept_lylty.sort_values(by = 'Loyal customer', ascending = True).plot(kind = 'b

# Adding Chart Labels
sns.set_style('ticks')
plt.ylabel('DEPARTMENT')
plt.xlabel('ORDER FREQUENCY (in millions)')
plt.xticks(fontsize = 10, fontweight = 'bold')
plt.yticks(fontsize = 10)
plt.title('LOYALTY STATUS ORDERING HABITS BY MOST POPULAR PRODUCTS', fontweight = 'bold')

plt.show()
```

FIG. 4f

VIEW FULL REPORTS BELOW

Python Codes | Final Excel Report

RECOMMENDATIONS

CONSUMER TREND

ADVERTISEMENT

Product endorsements of the least popular products like international, pantry, etc. should be promoted on weekends (Fridays, Saturdays, and Sundays) between 9 am and 3 pm, leveraging fast-moving goods such as dairy eggs, beverages, and beverages so on.

CUSTOMER SEGMENT

Offering special rates and rewards through a point system tailored to members' interests maintains loyal buyers and converts non-frequent buyers' status to a higher level.

COMPETITOR ANALYSIS

Looking into competitors' pricing schemes and conducting surveys on consumers' household expenditures can disclose key information for effective pricing tactics.

RESEARCH & DEVELOPMENT

Expand the services to new target audiences that may be interested in the premium brands and an exclusive offer on executive packages can spark new interests and eventually extend business in that area.

VIEW FULL REPORTS BELOW

Python Codes | Final Excel Report