Instacart Project Brief

Haley Bustle

November 23, 2023

**Project Title: Instacart Sales Study, 2023**

**Objective**: Conduct an analysis of Instacart’s sales and customer data to respond to questions regarding sales and future marketing strategies. Note: Instacart is a real company that has made their data available online. However, the contents of this project brief have been fabricated for the purpose of this Achievement.

**Project Context:**

* Understand data to develop targeted marketing strategies
* analyze customers in their behaviors and interactions with Instacart
* identify how different customer profiles impact marketing

**Key Questions to Address:**

1. Identify the busiest days and hours for scheduling ads.
2. Determine the times of day people spend the most.
3. Group products into simpler price range categories for targeted marketing.
4. Find the most popular types of products and the highest frequency of product orders by department.
5. Analyze customer brand loyalty, differences in ordering habits based on loyalty status, region, age, and family status.
6. Classify customers based on demographics, order frequency, and product preferences.

**Stakeholders:**

* Vice President of Marketing
* Senior Vice President of Sales
* Instacart Customers

**Data Sources:**

* Customers Data Set (provided by CareerFoundry)
* Instacart Data Sets (Data Dictionary, 2017 Online Grocery Shopping Dataset)

**Deliverables:**

* Analysis conducted using Python and relevant libraries.
* Recommendations for Instacart stakeholders based on analysis.
* Final report with evidence of methodology, analysis, and data citation.

**Terminology:**

● script = notebook

● variable = column = characteristic

● observation = entry

● dataset = dataframe = df

● read = import

● run = execute

● write = export = save

● derive a variable = create a column

● filter = subset

● merge flag = match flag

● key column = identifier column

**Dataframe Terminology:**

To allow an easier and cleaner presentation of visualizations, numbers are used in place of the days of the week:

0 = Saturday

1 = Sunday

2 = Monday

3 = Tuesday

4 = Wednesday

5 = Thursday

6 = Friday

**When understanding [the] hour\_of\_day column, this information is in military time. For example, 10 = 1000 hours or 10:00 just as 15 = 1500 hours or 3:00pm**

**Project Timeline:**

**Exercise 1: Intro to Programming for Data Analysts**

● Install Anaconda

● Launch Jupyter

**Exercise 2: Jupyter Fundamentals & Python Data Types**

● Create project folder

● Install required Python libraries

● Create a notebook and import libraries

● Practice coding using basic Python data types

**Exercise 3: Introduction to Pandas**

● Download data and import into notebook as a pandas dataframe

● Conduct basic descriptive exploratory tasks

**Exercise 4: Data Wrangling & Subsetting**

● Change data types of identifier variables into more suitable types and rename

columns where appropriate

● Access values and determine their meaning using a data dictionary

● Create new dataframes based on a certain criterion

● Answer questions about user activities based on variable frequencies

**Exercise 5: Data Consistency Checks**

● Fix mixed-type variables

● Uncover and deal with missing values

● Uncover and remove duplicates

**Exercise 6: Combining & Exporting Data**

● Merge a set of given dataframes

● Analyze results from merge flag frequencies

● Export merged data as a pickle file

**Exercise 7: Deriving New Variables**

● Create new columns using conditional logic in the form of if-statements, user-defined

functions, the loc() function, and for-loops

**Exercise 8: Grouping & Aggregating Data**

● Create flags, for instance, a loyalty flag, and place them in new columns

● Create summary columns of descriptive statistics using the groupby() function

**Exercise 9: Intro to Data Visualization with Python**

● Import and prepare a customer data set

● Merge customer data with other project data

● Create histograms, bar charts, line charts, and scatterplots for different variables and

relationships between variables

**Exercise 10: Coding Etiquette & Excel Reporting**

● Create new columns and flags using customer data to inform customer profiling

● Analyze order behavior of different customer groups

● Summarize analysis findings and describe what connections in the data you’ve found

● Create a report that describes your analysis methodology, your results, and your

recommendations for Instacart stakeholders