CAREERFOUNDRY

Achievement 4 Project Brief: Instacart Grocery Basket Analysis

Note: Instacart is a real company that's made their data available online. However, the contents of this project brief have been fabricated for the purpose of this Achievement.

Objective

You're an analyst for an existing company, Instacart, an online grocery store that operates through an app. Instacart already has very good sales, but they want to uncover more information about their sales patterns. Your task is to perform an initial data and exploratory analysis of some of their data in order to derive insights and suggest strategies for better segmentation based on the provided criteria.

Context

The Instacart stakeholders are most interested in the variety of customers in their database along with their purchasing behaviors. They assume they can't target everyone using the same methods, and they're considering a targeted marketing strategy. They want to target different customers with applicable marketing campaigns to see whether they have an effect on the sale of their products. Your analysis will inform what this strategy might look like to ensure Instacart targets the right customer profiles with the appropriate products. The stakeholders would like to be able to answer the following key questions:

Key Questions

- The sales team needs to know what the busiest days of the week and hours of the day are (i.e., the days and times with the most orders) in order to schedule ads at times when there are fewer orders.
- They also want to know whether there are particular times of the day when people spend the most money, as this might inform the type of products they advertise at 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 of products that are more popular than others? The marketing and sales teams want to know which departments have the highest frequency of product orders.

Note: Instacart is a real company that's made their data available online. However, the contents of this project brief have been fabricated for the purpose of this Achievement.

- The marketing and sales teams are particularly interested in the different types of customers in their system and how their ordering behaviors differ. For example:
    - What's the distribution among users in regards to their brand loyalty (i.e., how often do they return to Instacart)?
    - Are there differences in ordering habits based on a customer's loyalty status?
    - Are there differences in ordering habits based on a customer's region?
    - Is there a connection between age and family status in terms of ordering habits?
    - What different classifications does the demographic information suggest? Age? Income? Certain types of goods? Family status?
    - What differences can you find in ordering habits of different customer profiles? Consider the price of orders, the frequency of orders, the products customers are ordering, and anything else you can think of.

Stakeholders

- **Vice President of Marketing:** "We're always looking into improving our targeting for ad campaigns."

- **Senior Vice President of Sales:** "We need to know what part of our offering has the lowest market share and why. Based on this input, we could improve this sector and boost sales."

- **Instacart Customer:** "I want to receive ads, promotions, and recommendations that are relevant to the products I order regularly."

Data

Throughout this Achievement, you'll be using a number of open-source data sets from Instacart. You'll also receive a customer data set (created and included for the purpose of this project), on which you'll apply what you've learned to address the project's key questions. While each data set contains a different kind of information, they all include some kind of common identifier.

Note: Instacart is a real company that's made their data available online. However, the contents of this project brief have been fabricated for the purpose of this Achievement.

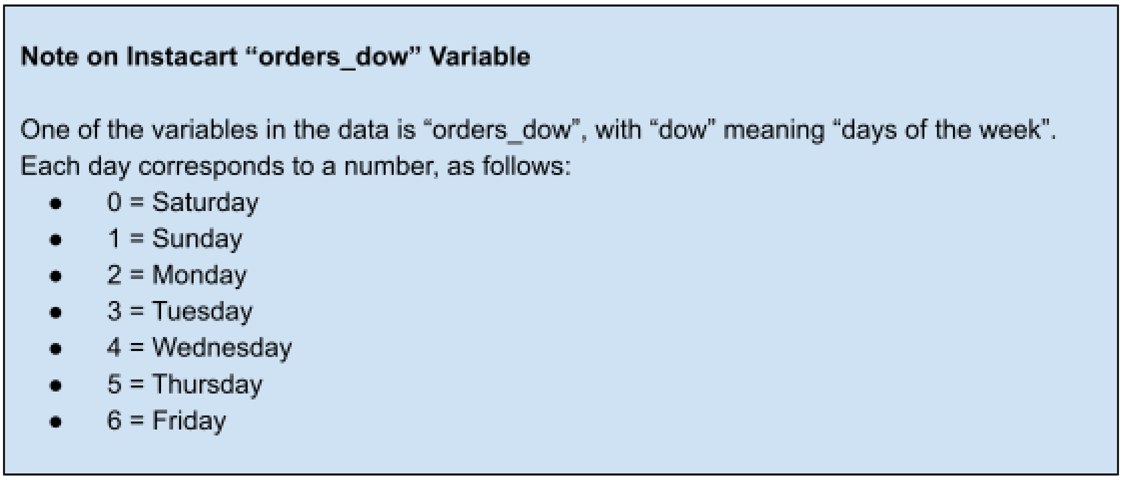
The project data you'll need is linked for reference below. However, you'll receive links to each data set in the Exercise content, as well.

CareerFoundry Data Sets:

- Customers Data Set

Instacart Data Sets:

- Data Sets
- Data Dictionary
- Citation (required in your final report): "The Instacart Online Grocery Shopping Dataset 2017", Accessed from https://www.instacart.com/datasets/grocery-shopping-2017 on <date>.



Note on Instacart "orders_dow" Variable

One of the variables in the data is "orders_dow", with "dow" meaning "days of the week". Each day corresponds to a number, as follows:

- 0 = Saturday
- 1 = Sunday
- 2 = Monday
- 3 = Tuesday
- 4 = Wednesday
- 5 = Thursday
- 6 = Friday

Analysis Criteria

- Project folder follows industry standards in terms of structure and naming conventions.
- Analysis has been conducted using Jupyter notebooks and the Anaconda libraries manager.
- Analysis has been conducted using Python and relevant libraries (pandas, NumPy, os, matplotlib, scipy, and seaborn).

Note: Instacart is a real company that's made their data available online. However, the contents of this project brief have been fabricated for the purpose of this Achievement.

- All required libraries have been successfully installed and imported into each script.
- Python scripts are clean and easy to follow with headings and contents lists.
- All code is consistent (e.g., with the use of quotation marks and spaces) and includes descriptive comments.
- All required data sets have been successfully installed and imported into each script.
- Descriptive checks have been conducted after importation of data, such as checking the top and the bottom of the dataframe.
- Whenever a dataframe is altered, checks have been conducted to determine its shape and basic statistics.
- All project data has been merged into a single data set. A frequency of the merge flag shows the merged data set is a 100% match to the combined original data sets.
- Merged data set only contains variables to be used in the analysis.
- All column names are self-explanatory.
- All identifier variables follow the industry standard data type.
- Data has been cleaned. Duplicate data, missing data, and mixed-type columns have been checked and addressed.
- Samples have been exported whenever an exclusions flag has been created.
- All subsamples have been exported and saved in the proper folder following a consistent naming convention.
- Any new columns that have been derived are relevant to the needs of the analysis.
- At least 4 types of data visualizations have been generated to communicate insights to stakeholders. Visualizations are clearly labeled.
- Data ethics have been kept in mind when dealing with data, especially in regards to customer information.
- Final report includes evidence of analysis methodology, clear answers to the questions in this brief, and recommendations for Instacart stakeholders.
- Final report contains data citation for Instacart and customer data sets.

Terminology

In analytics, a single procedure or concept can often be called a variety of different things. We've aimed to be consistent with the terminology used throughout this Achievement. Even so, there are a few variations that come up again and again when conducting an analysis in Python. We've included a list below to help you navigate this terminology:

Note: Instacart is a real company that's made their data available online. However, the contents of this project brief have been fabricated for the purpose of this Achievement.

- 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

Project Tasks & Deliverables

Throughout this Achievement, you'll work on your project from one Exercise to the next, completing tasks as you go. For each task, you'll submit a deliverable that makes up a piece of your project. Below is a breakdown of your tasks and deliverables by Exercise:

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 needed
- Access values and determine their meaning using a data dictionary
- Create new dataframes based on a certain criteria
- 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

Note: Instacart is a real company that's made their data available online. However, the contents of this project brief have been fabricated for the purpose of this Achievement.

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

Note: Instacart is a real company that's made their data available online. However, the contents of this project brief have been fabricated for the purpose of this Achievement.