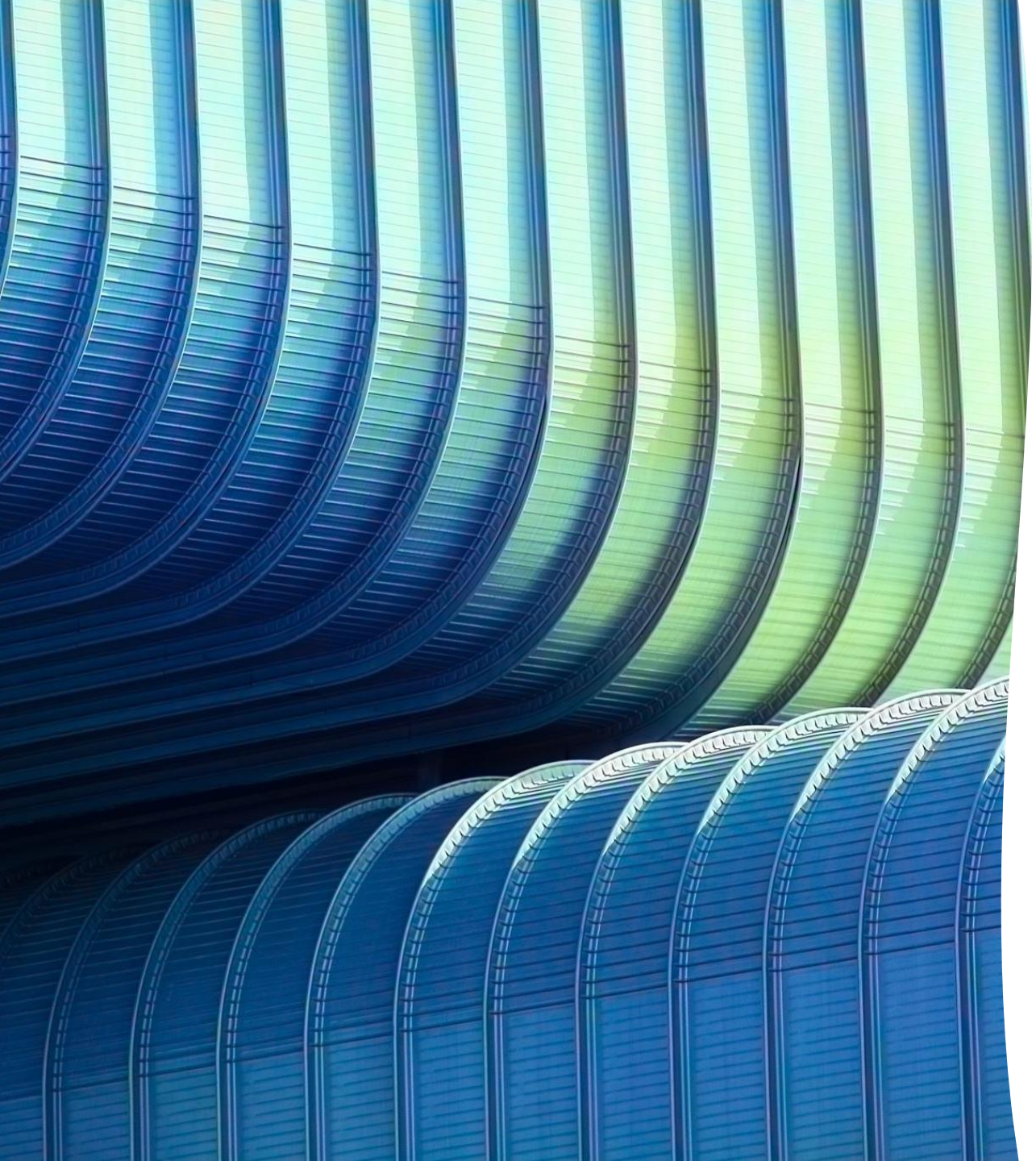

Trends and Analysis
of eCommerce Data



Group Team Members

Francesca Barrios

Laura Brown

Seth Ely

Darby Hansen

Description

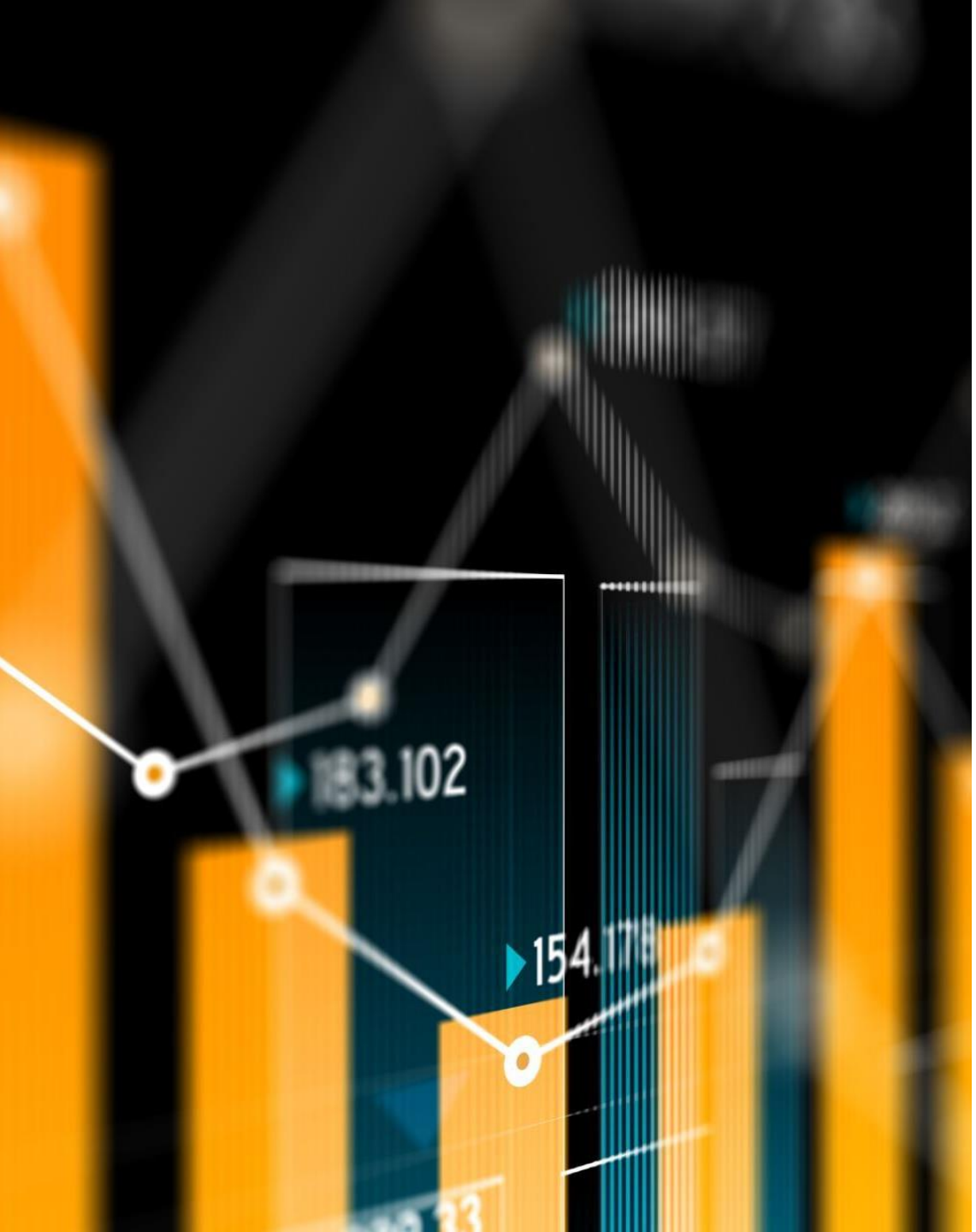
Our objective is to gain valuable insights into the trends across various brands and categories within the realm of online eCommerce retail, mirroring real-world marketing scenarios.

To achieve this, we will analyze the behavioral data sourced from a substantial multi-category online store, compiled by the Open CDP project spanning October 2019 to April 2020.

Which brands, items, or item categories sell the most product?

Which brands, items, or item categories have the highest purchases per view?

What are some seasonal trends we can determine over the 7 month window of this dataset?



Prior Work

- Numerous studies are dedicated to uncovering trends in eCommerce, reflecting the dynamic nature of the industry.
- For example, "Global eCommerce Market Analysis & Trends" by Grand View Research offers a comprehensive analysis of the global eCommerce landscape.
- This study delves into growth projections, key drivers, challenges, and emerging trends shaping the eCommerce market.

- Grand View Research. (2020). "Global eCommerce Market Analysis & Trends." Retrieved from [https://www.grandviewresearch.com/industry-analysis/e-commerce-market].

Datasets

E-Commerce user Behavior

Data compiled by the OpenCDP Project and available form Kaggle:

https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store

While It is possible to import this dataset into a Pandas dataframe directly from the URL, the load time is substantial due to the size of each monthly dataset.

For this reason, we will be downloading the dataset individually and uploading our findings to our github repository located at

https://github.com/seel6470/CSPB-4502-Group-9-Project

October 2019	https://data.rees46.com/datasets/marketplace/2019-Oct.csv.gz
November 2019	https://data.rees46.com/datasets/marketplace/2019-Nov.csv.gz
December 2019	https://data.rees46.com/datasets/marketplace/2019-Dec.csv.gz
January 2020	https://data.rees46.com/datasets/marketplace/2020-Jan.csv.gz
February 2020	https://data.rees46.com/datasets/marketplace/2020-Feb.csv.gz
March 2020	https://data.rees46.com/datasets/marketplace/2020-Mar.csv.gz
April 2020	https://data.rees46.com/datasets/marketplace/2020-Apr.csv.gz

Proposed Work

Data Cleaning:

- Use inference-based process to fill in missing values in "category_code" based on brand names.
- Remove duplicate entries, as to not skew the data.

Data Reduction:

- Remove irrelevant attributes that will not be used to answer our overall questions, such as event_time, user_id, and user_session.

Data Transformation:

- Use different groupby functions to group the different brands, items, and item categories together to create a more condensed table.

List of tools

The tools we anticipate using during our evaluation of this dataset are mostly Python libraries:

- Seaborn
- Pandas
- Numpy
- Matplotlib

Additionally, we may use Tableau. This tool would help generate different visualizations of our findings, potentially offering a better fit than what can be generated using the Python libraries.

Evaluation

- **Accuracy**
  - Based on our findings, we can train a model and interpret the validity of the results using the accuracy, precision, and recall metrics.
- **Relevance**
  - We will evaluate our findings to see how relevant, applicable, or interesting they would be by comparing them to the anticipated needs of real-world eCommerce retail corporations.
- **Interpretability**
  - We will improve the interpretability of our findings by using charts, graphs, and visualizations. Additionally, we will use feature selection to identify the features that present the strongest correlation to create a compelling narrative from our insights.