FAVORITA STORE SALES PREDICTION

OVERVIEW

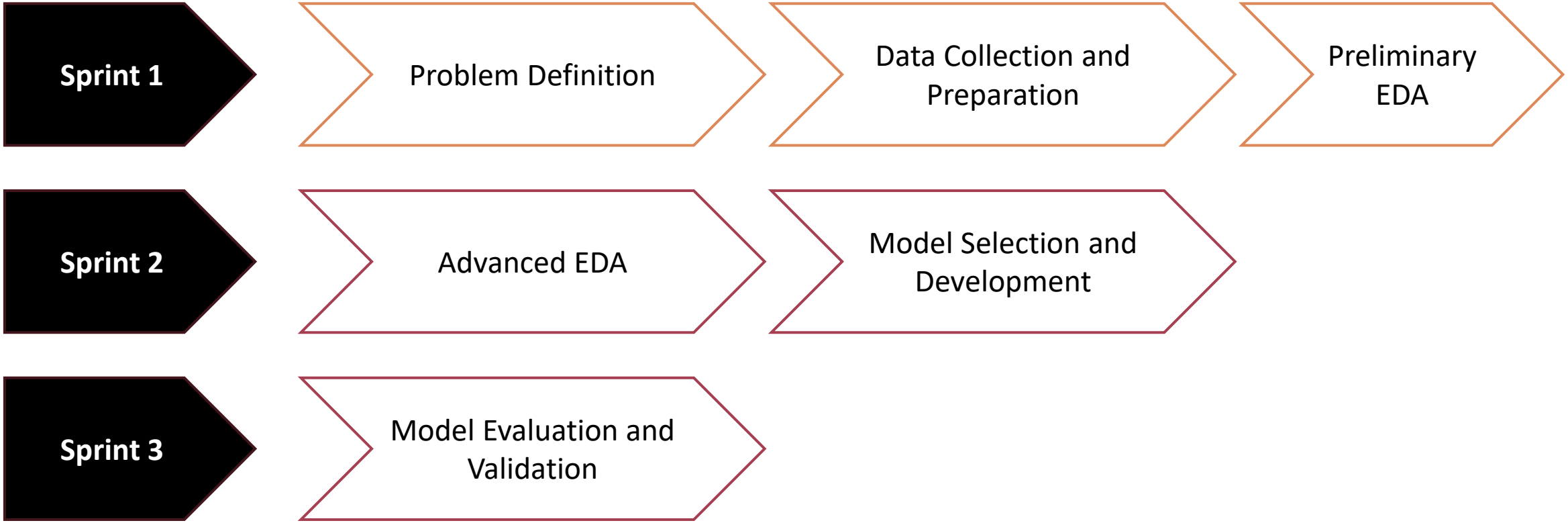
About the Business

- Favorita is one of the **largest supermarket chains** in **Ecuador**, known for its extensive selection of **groceries, household items, and other goods**. It has **numerous locations** across the country.

Purpose of the Project

- Give past sales data of different categories and stores, **predict future daily sales of each category in different stores**.
- Prevent **overstocking** and **understocking**
- Improve **customer experience**
- **Control business cost**

PROJECT CADENCE



IMPACT OF THE PROJECT

Lower Costs

Improve Customer Satisfaction & Business Efficiency

ABOUT THE DATASET

- ***5*** tables: ***3*** million rows and ***21*** columns, no duplicates, ***43 null*** values in the oil price table to be filled manually
- Category daily sales units between 2013 and 2017, promotions, oil price, holidays and festivals, store locations.
- Top10 families by sales units: 'GROCERY I', 'BEVERAGES', 'PRODUCE', 'CLEANING', 'DAIRY', 'BREAD/BAKERY', 'POULTRY', 'MEATS', 'PERSONAL CARE', 'DELI’
- to predict ***daily sales*** for ***33 categories*** across ***54 stores*** in Ecuador.

MY QUESTIONS

- After using a time series model, should I compare the predictions with those from other models?
- What are the general steps for the project? What are the timelines for each procedure? Do I need to build a GUI to take data and make predictions in real-time?
- For my data, if I want to predict daily sales for each family in different stores in different cities, do I need to convert all the text into numeric values before running a time series model? (There are many cities and families that are in text format)?


THANK YOU