INSTACART

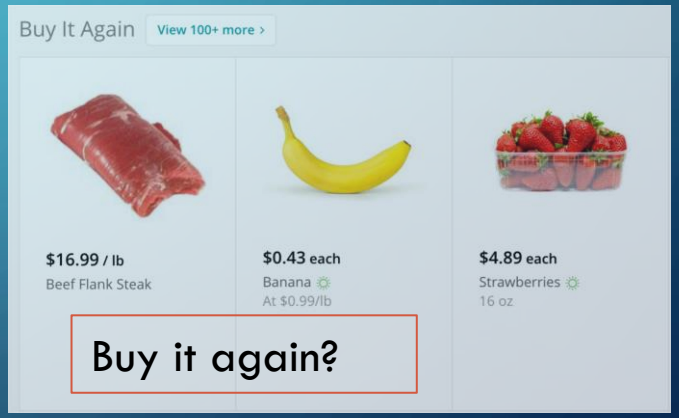
PREDICTING NEXT SHOPPING CART

Dmitri Gourianov

OVERVIEW

Motivation: Understanding purchasing patterns of a consumer can lead to increased sales and improved customer satisfaction

Objective: To create a classification model to predict which products will be in the user's next order

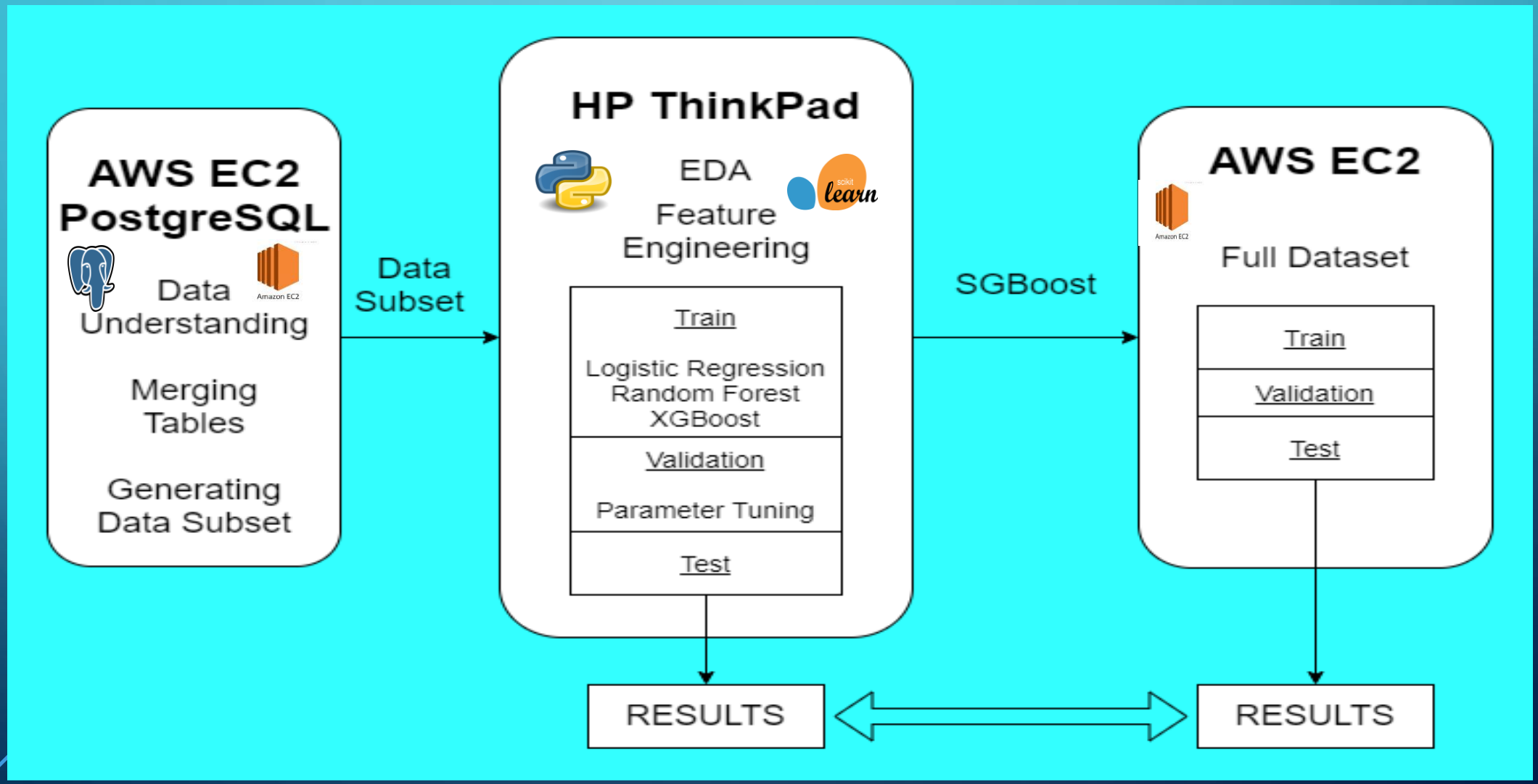


DATASET

Instacart database includes:

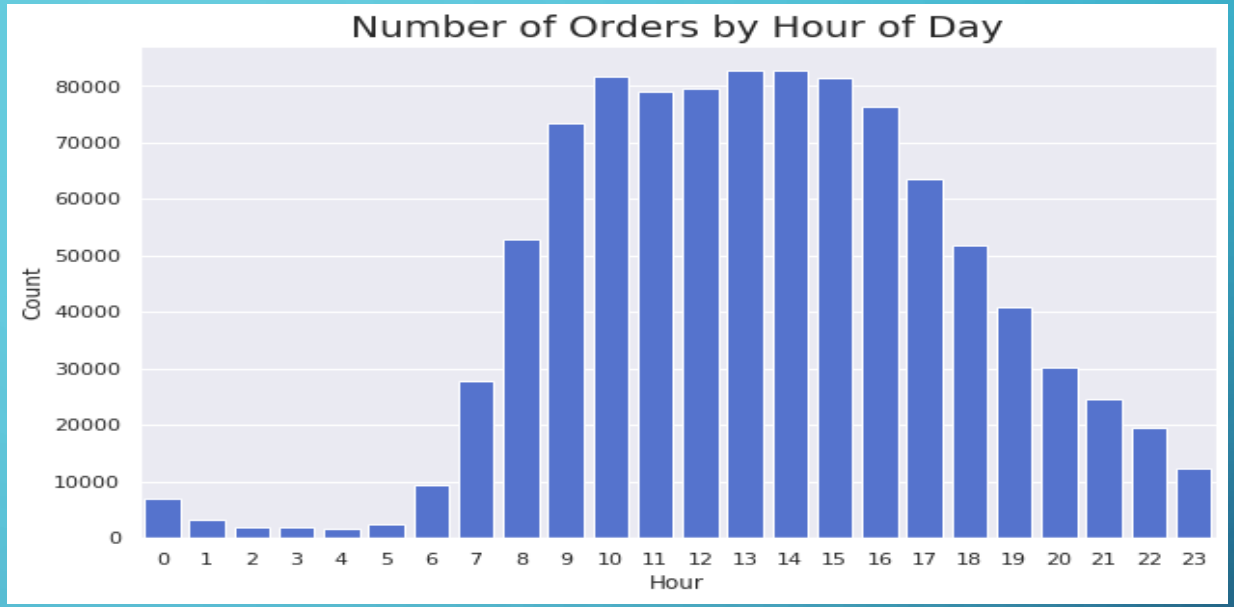
- 6 relational tables
- 3,000,000 grocery orders
- 200,000 users
- 4 to 100 orders per user
- 50,000 products
- Aisles and departments information

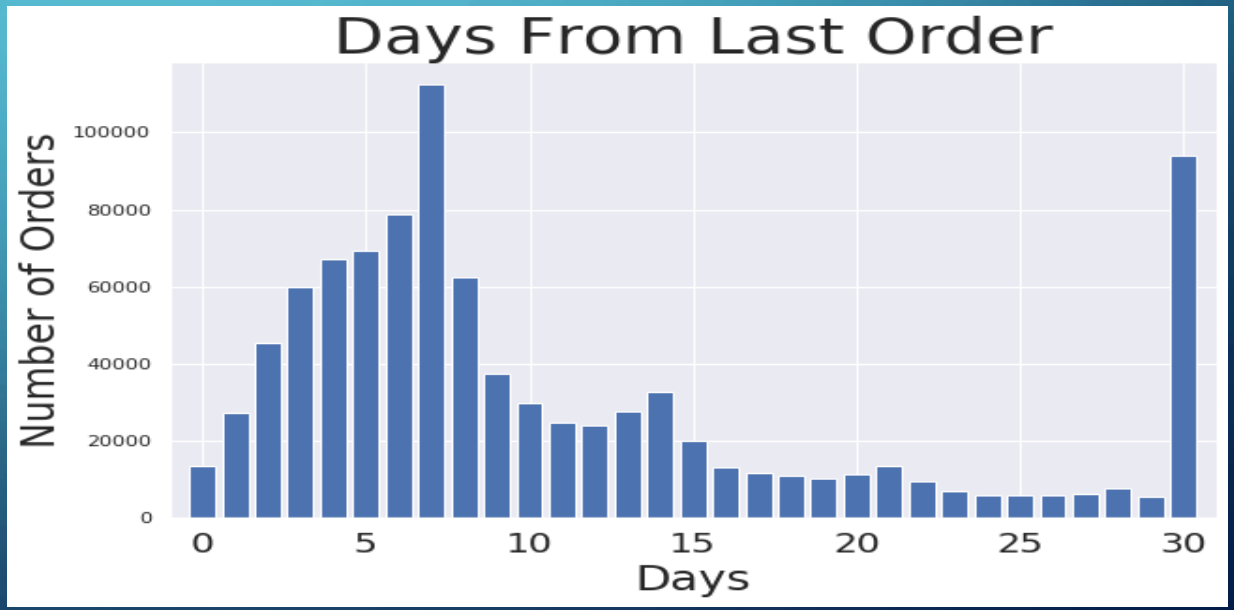
MACHINE LEARNING PIPELINE



EDA

- 8PM – Ice Cream time?
- Majority of the customers do their purchases weekly and monthly





FEATURE ENGINEERING

Product Features

User Features

Product-User Features

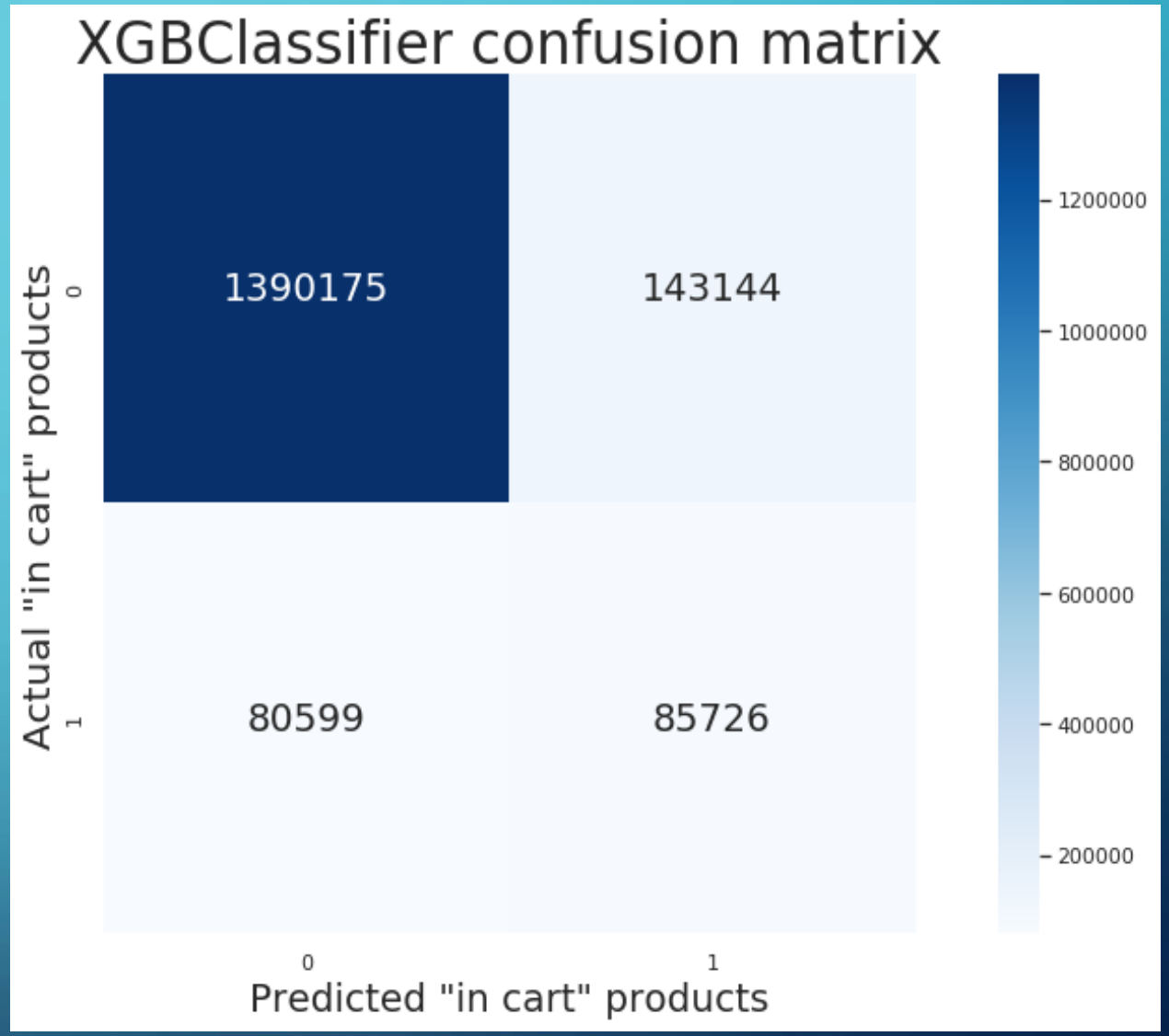
Aisle-Department Features

MODELS EVALUATION

	Baseline Models		Logistic Regression	Random Forest	XGBoost
	All Products	Last Order			
F1	**0.17**	**0.17**	**0.40**	**0.41**	**0.42**
Precision	0.09	0.15	0.33	0.42	0.37
Recall	1.0	0.19	0.50	0.40	0.47

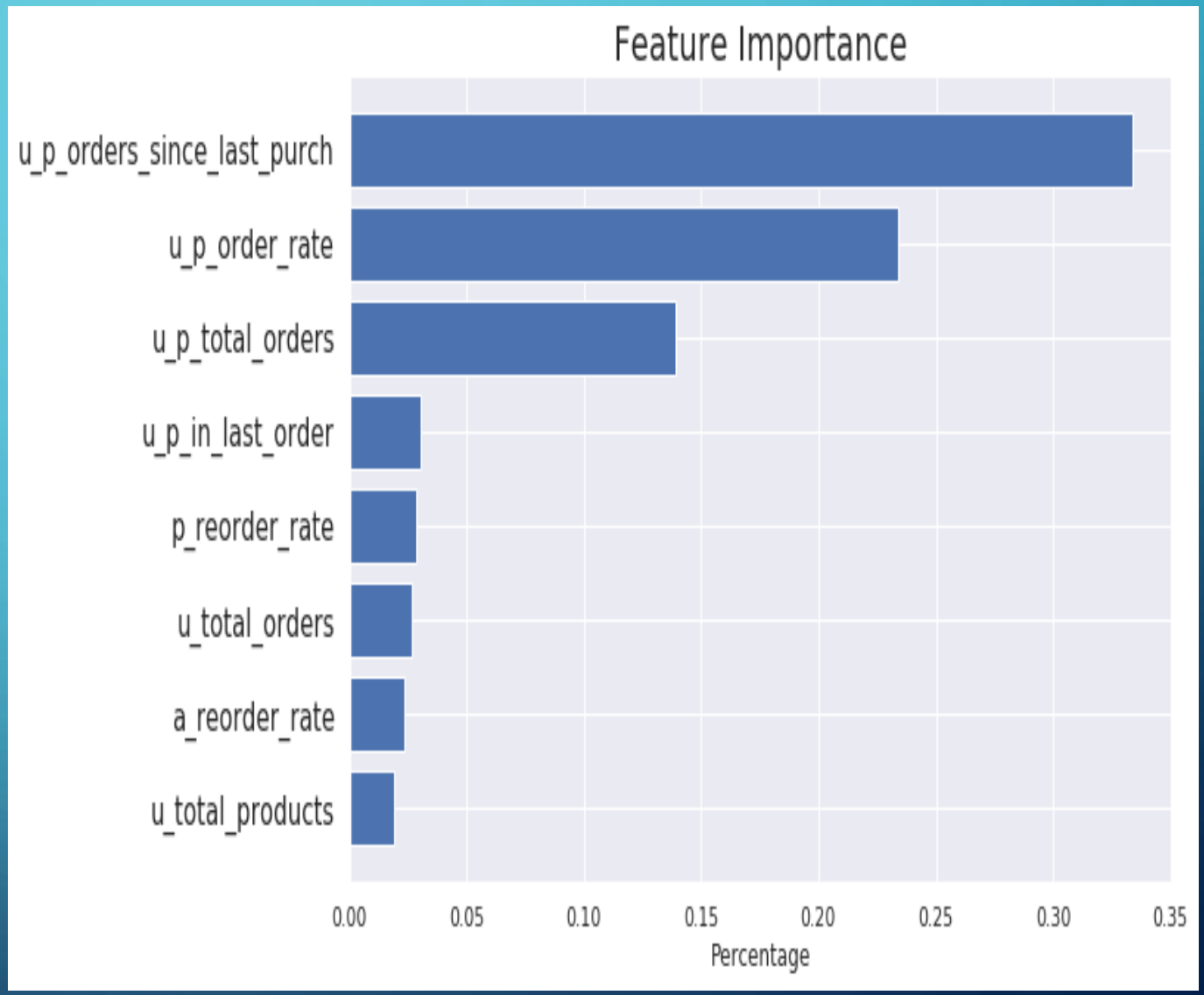
FULL DATASET RESULTS ON AWS

Metrics	Score
F1 Score	0.43
Precision	0.37
Recall	0.51



FEATURE IMPORTANCE

- Most important features are related to product-user interaction
- Top three features account for more than 75% of the 'importance'



PRACTICAL APPLICATIONS OF THE MODEL:

- Reminding the customers to reorder the products
- Products can be added to the cart automatically based on the customer preferences
- Targeted promotional and marketing campaigns
- Customers can be given additional offers by bundling the products together
- Instacart can use the model to ensure the replenishments and proper scheduling of the products

FUTURE WORK

- Improve model by collecting additional data, such as seasonality, user's location or gender
- Identify the clusters and subgroups of customers
- Investigate additional consumer behavior topics such as customer churn

THANK YOU !!!

APPENDIX