

Fin
Techs

Capstone Project

Google Analytics Customer Revenue



FinTechs presenters:
Hassan Salam, Jonathan Preiss, Sina Lotfiomran, Davain Pablo Edwards

27.4.2021

The Team: FinTechs

Hassan Salam

- ❖ Physics Researcher
- ❖ Electronics Eng.
- ❖ Sales Analyst



Jonathan Preiss

- ❖ Online Marketing
- ❖ Business Development
- ❖ Data Scientist



Sina Lotfimran

- ❖ Mechanical Engineer
- ❖ Research Assistant
- ❖ DS with Engineering



Davain Pablo Edwards

- ❖ Experienced Full Stack Developer
- ❖ Trained Computer Science Expert
- ❖ Data Scientist



Agenda

- ❖ Business Case
- ❖ Objectives & Approach
- ❖ Our Workflow
- ❖ Data Description
- ❖ Metric
- ❖ Exploratory Data Analysis
- ❖ Modeling
- ❖ Future Work

Business Case



Business Case

Stakeholder:

Publishers who wish to promote their products on the google merchandise store.

Goal:

Analyse the publishers online campaigns by given them insights for USA market, which will be able to predict if a customer will made a purchase or not and how much revenue the customer will generated.

Agenda

- ❖ Business Case
- ❖ Objectives & Approach
- ❖ Our Workflow
- ❖ Data Description
- ❖ Metric
- ❖ Exploratory Data Analysis
- ❖ Modeling
- ❖ Future Work

Objective

The goal was to analyze the customers total revenues for the period of **May 1st 2018** to **October 15th 2018** to predict the revenues on future customer behaviors.

Our Approach

- First we trained our data based on the **Sessions** features
- Then, we **predicted** the revenue per session
- Finally, We **aggregated** revenue of sessions per Customer.

Agenda

- ❖ Business Case
- ❖ Objectives & Approach
- ❖ Our Workflow
- ❖ Data Description
- ❖ Metric
- ❖ Exploratory Data Analysis
- ❖ Modeling
- ❖ Future Work

Our Workflow:



Cleaning

Removing unwanted data.



Preprocessing

Preparing the data.



Feature Engineering

Calculating and creating new features.



Regression

Creating the right model.



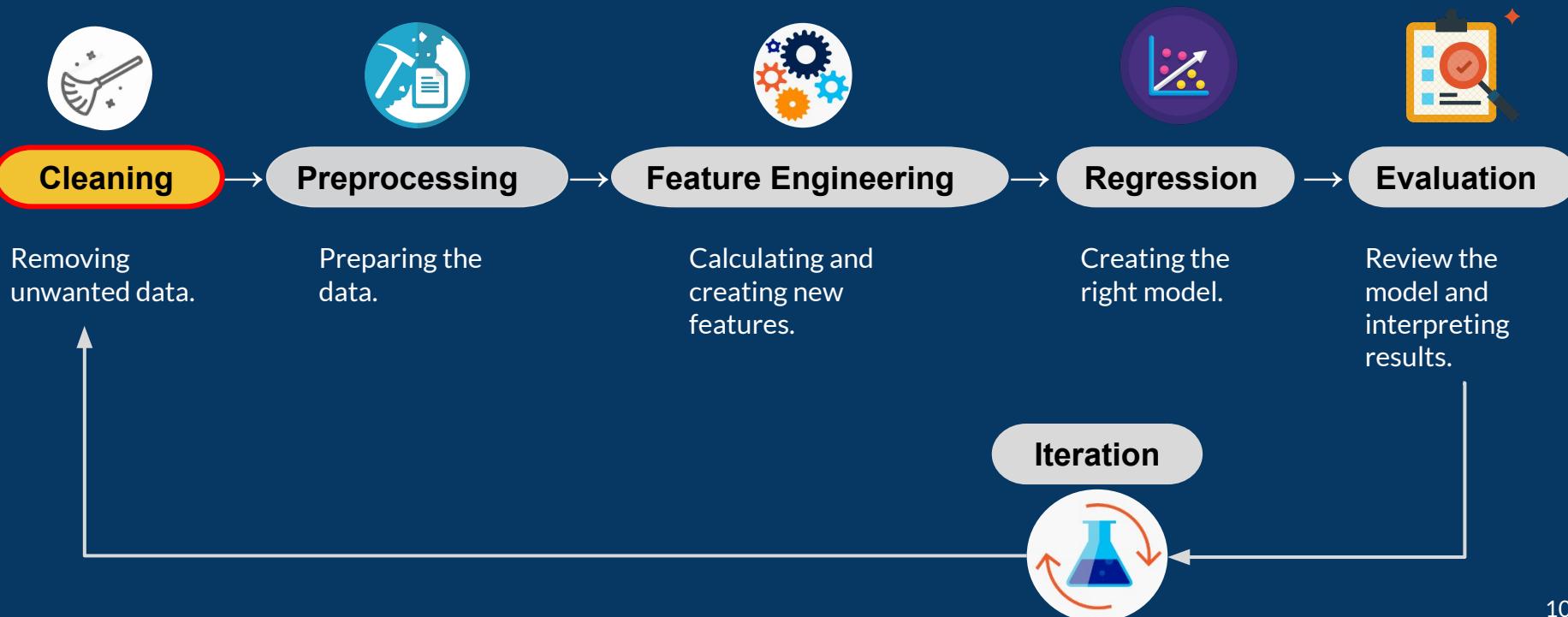
Evaluation

Review the model and interpreting results.

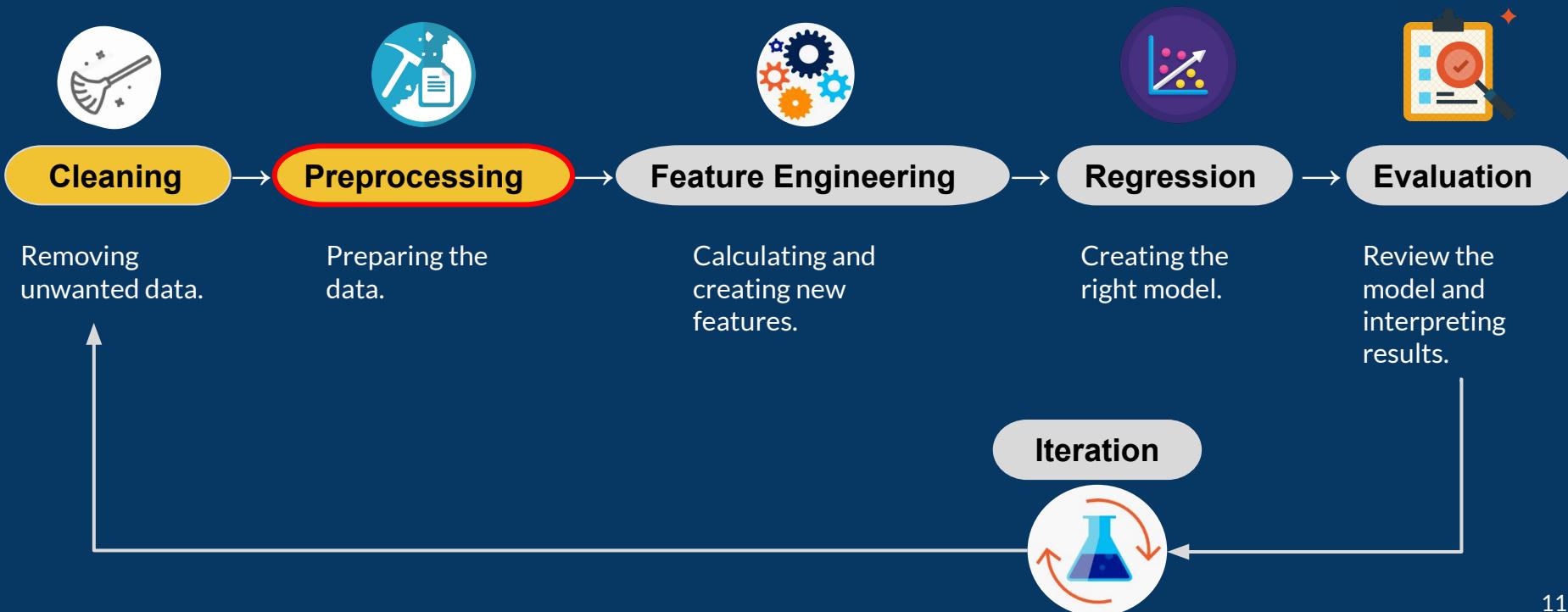
Iteration



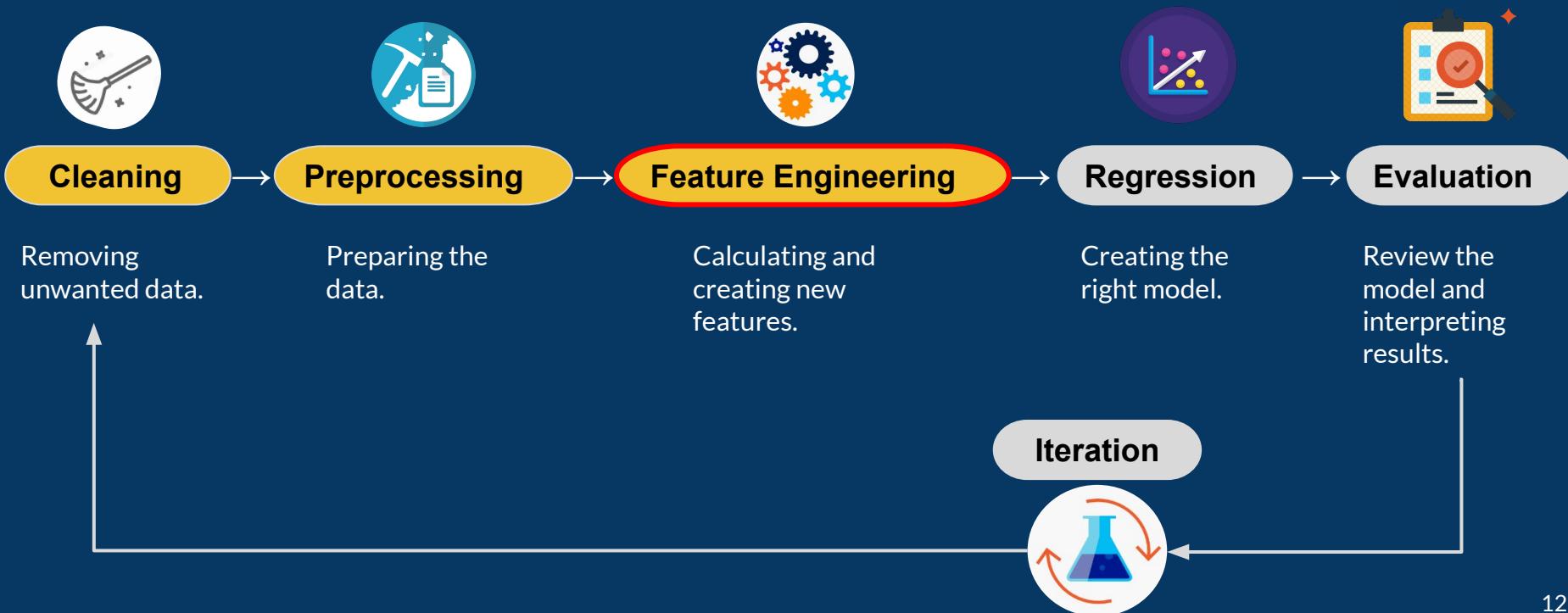
Our Workflow:



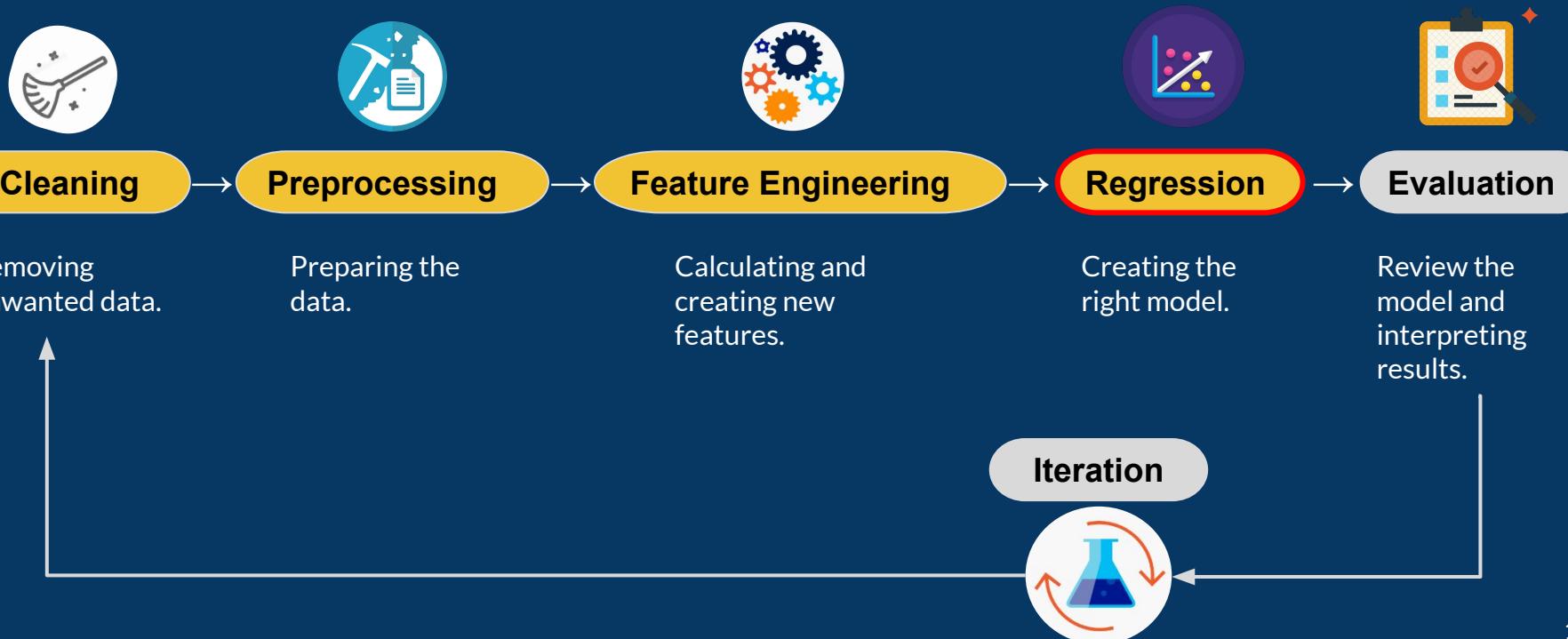
Our Workflow:



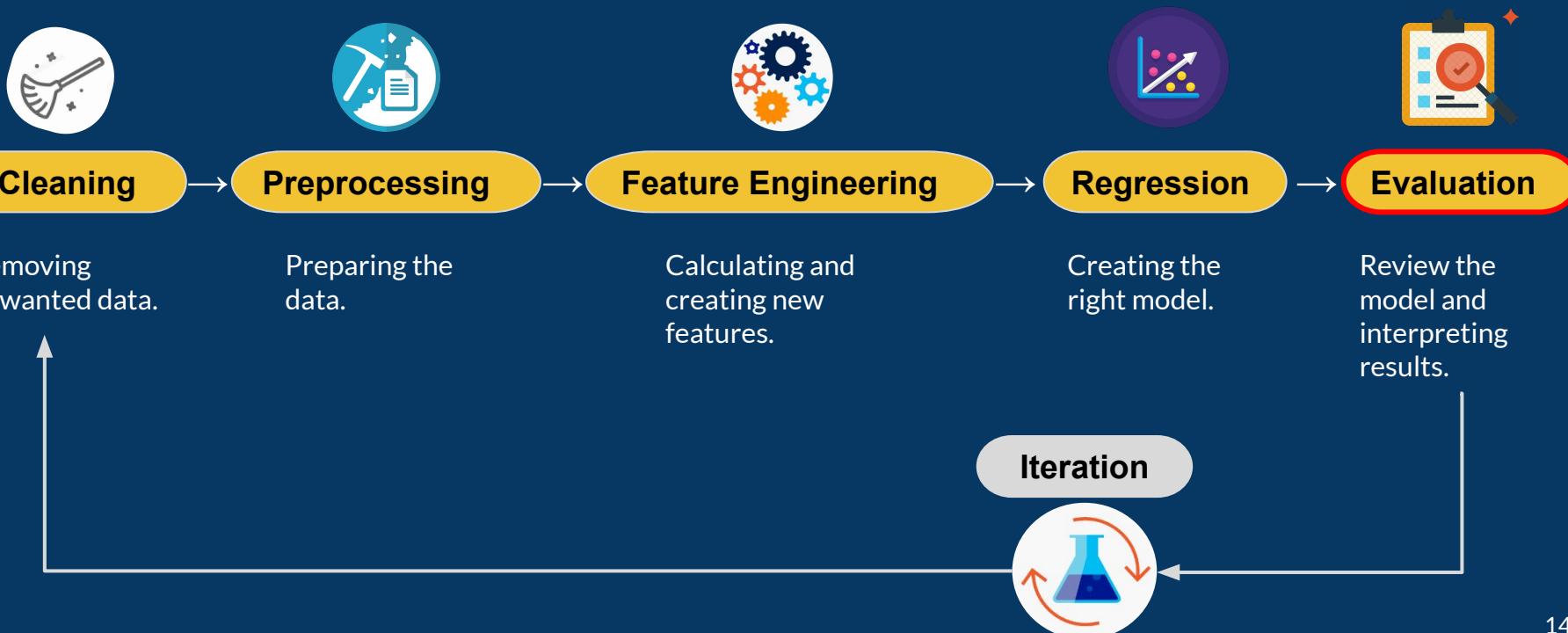
Our Workflow:



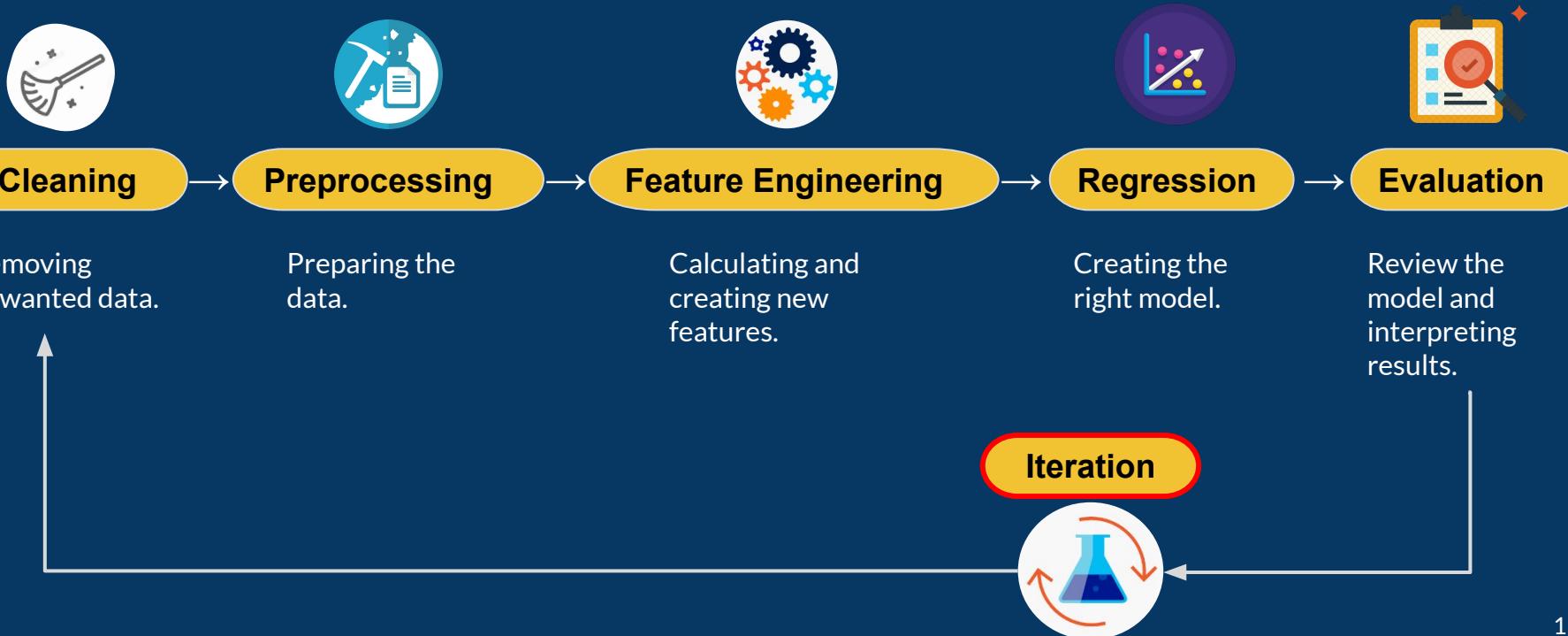
Our Workflow:



Our Workflow:



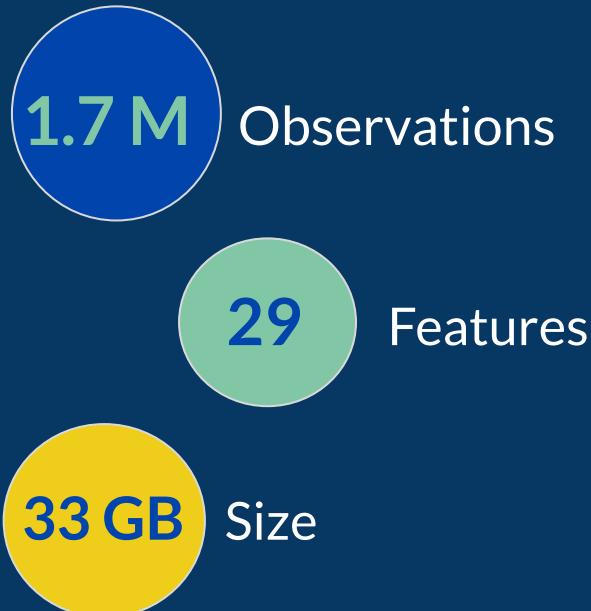
Our Workflow:



Agenda

- ❖ Business Case
- ❖ Objectives & Approach
- ❖ Our Workflow
- ❖ Data Description
- ❖ Metric
- ❖ Exploratory Data Analysis
- ❖ Modeling
- ❖ Future Work

Data Overview:



Some Features:

- Total Page View
- Total hits
- Week
- Visit Number
- Geonetwork - Country
- Geonetwork - Domain
- Geonetwork - City
- Geonetwork - Region
- Channel Grouping
- Device operating System

80% - 20% Rule In Business

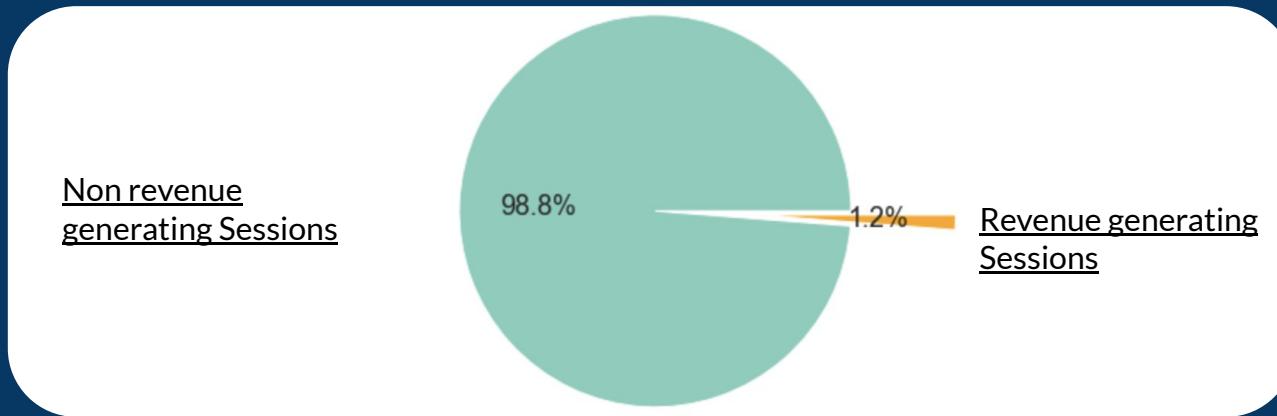


- The 80-20 rule, is a rule which asserts that 80% of outcomes, result from 20% of all causes.
- In business, a goal of the 80-20 rule is to identify inputs that are potentially the most productive and make them the priority.
- **In E-commerce businesses, it is more crucial to know the behaviour of your Sessions...**

... even more at Corona time!

Revenue Transactions Distributions

The distribution of revenue among sessions:



- ★ In our e-business it seems like 99% - 1% Rule
- ★ The struggle is real!

Agenda

- ❖ Business Case
- ❖ Objectives & Approach
- ❖ Our Workflow
- ❖ Data Description
- ❖ Metric
- ❖ Exploratory Data Analysis
- ❖ Modeling
- ❖ Future Work

Metric

- „totals.transactionRevenue“ Measure of predicting performance.

Root Mean Squared Error:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

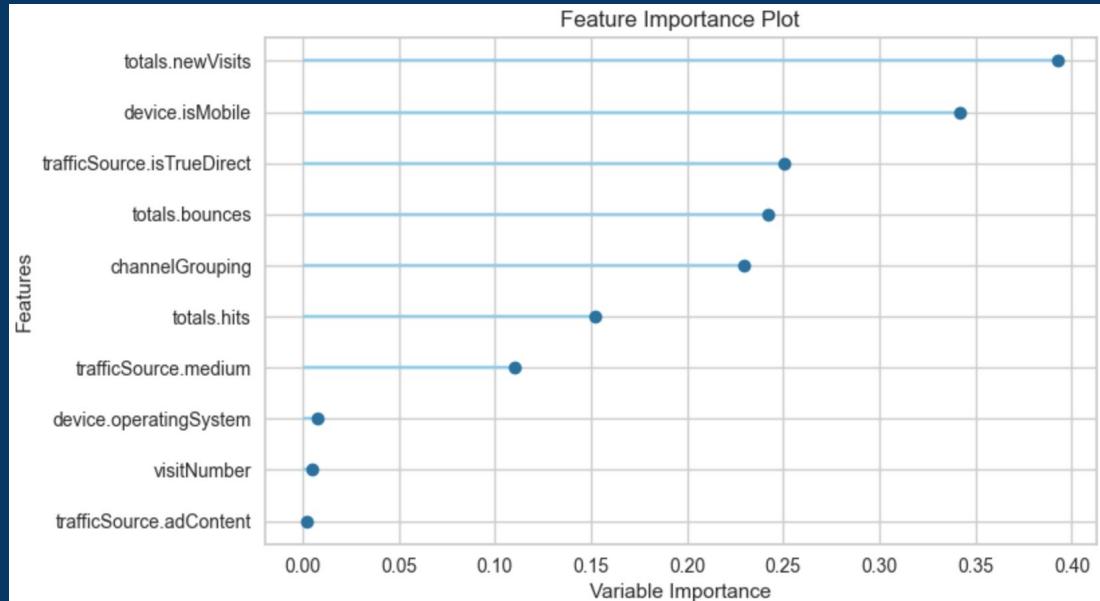
- \hat{y} (predicted-values): Is the natural log of the predicted revenue for a customer.
- y (actual-values): Is the natural log of the actual summed revenue value plus one.

Baseline Model

- All feature included
- No feature engineering
- Use lineare regression
- No hyper tuning
- No cross validation

Outcome:

RMSE: 2.05



Agenda

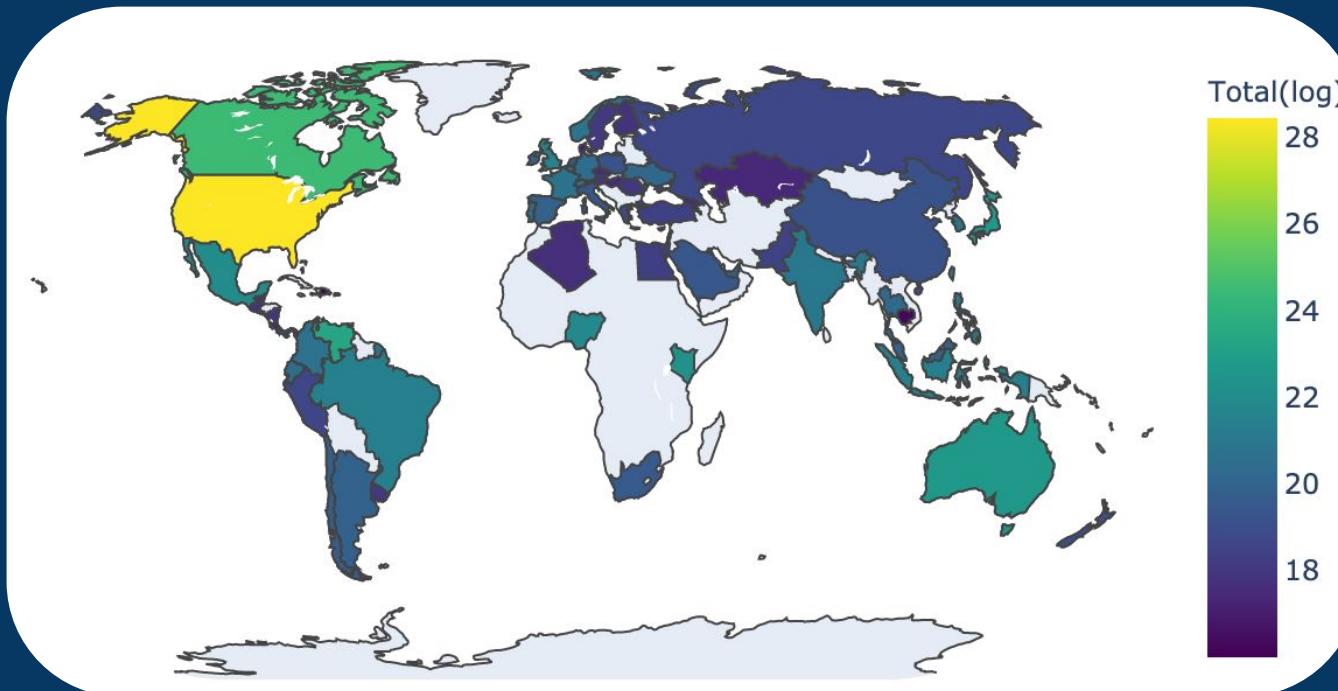
- ❖ Business Case
- ❖ Objectives & Approach
- ❖ Our Workflow
- ❖ Data Description
- ❖ Metric
- ❖ Exploratory Data Analysis
- ❖ Modeling
- ❖ Future Work

Keywords Customers type



- Most searches words are:
 - google
 - youtube
 - store
 - t-shirt

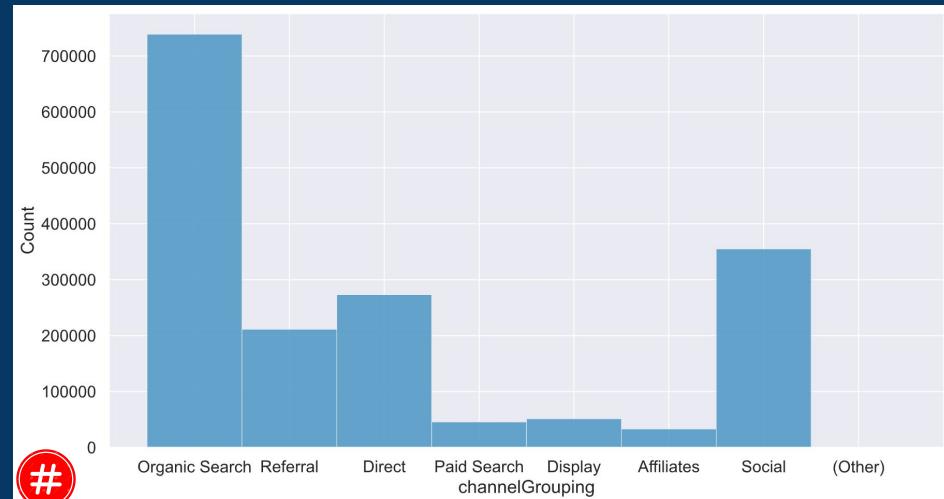
Total Revenue Transactions Per Country



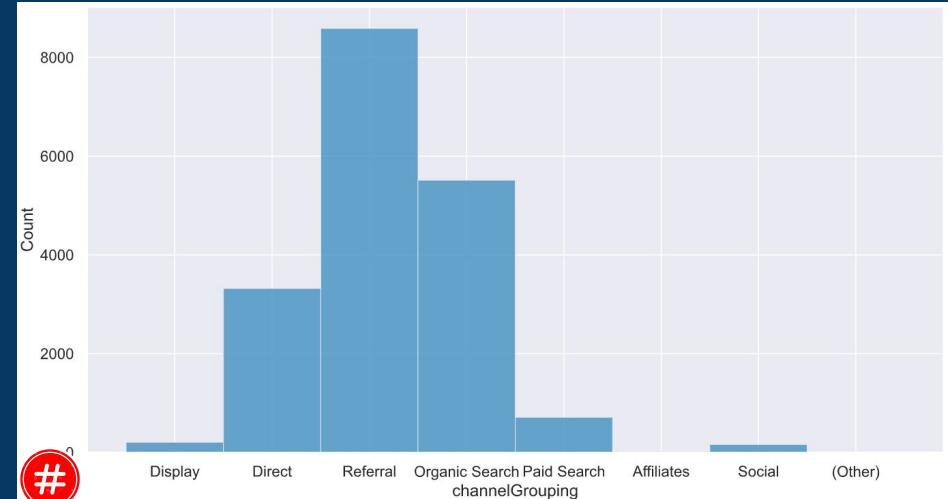
- The majority of the revenue is generated from the Americas.
- Few revenue generate from Asia.

1. Sessions coming from "Organic search" generate more Revenue

Channel distribution for Total Sessions:



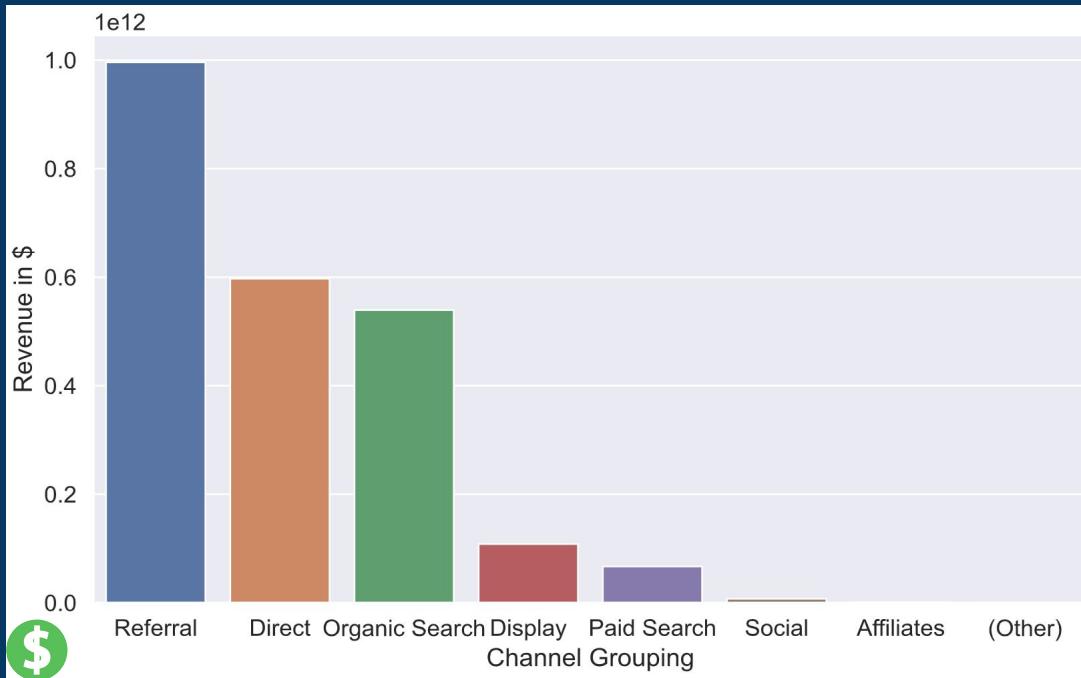
Channel distribution for Revenue Sessions:



- Total sessions includes Revenue + Non Revenue Sessions and come through Organic Search.
- However, the number of revenue generating Sessions from Referral is the highest

1. Sessions coming from "Organic search" generate more Revenue

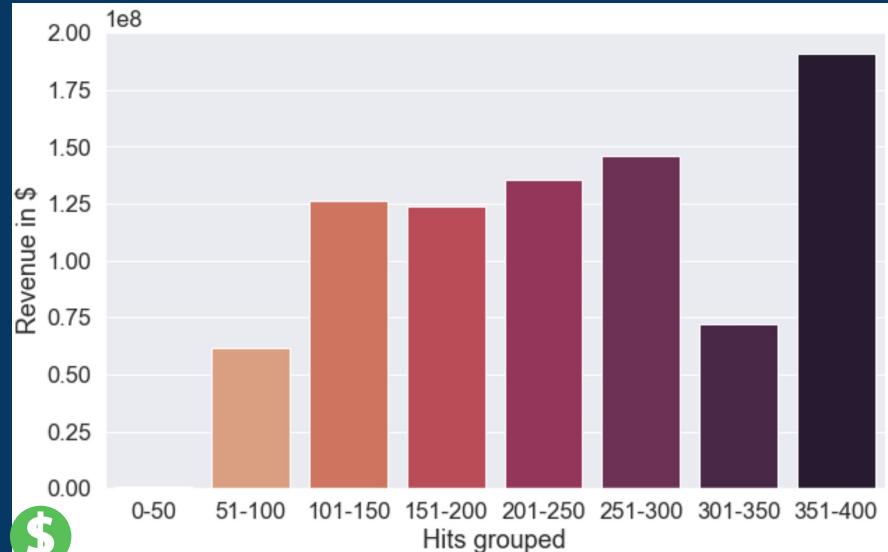
Revenue Transactions per channel:



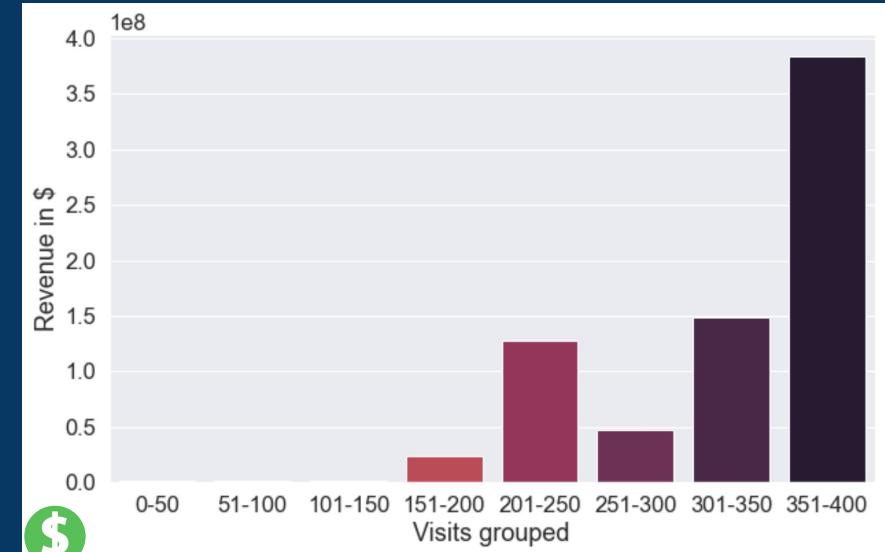
- **Referral** generate more Revenue even though the number of Sessions from Organic-Search is higher

2. Sessions with higher store visits made more Revenue on average

The average Revenue per number of hits group:



The average Revenue per number of visits group:

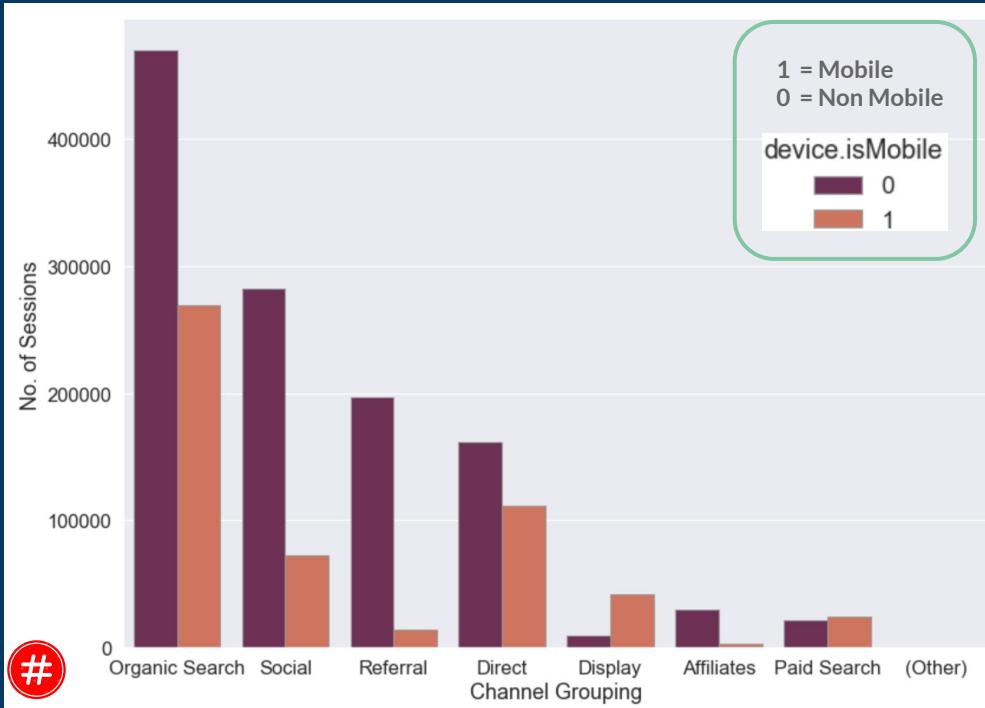


- The average revenue is generated from more number of hits and visits.

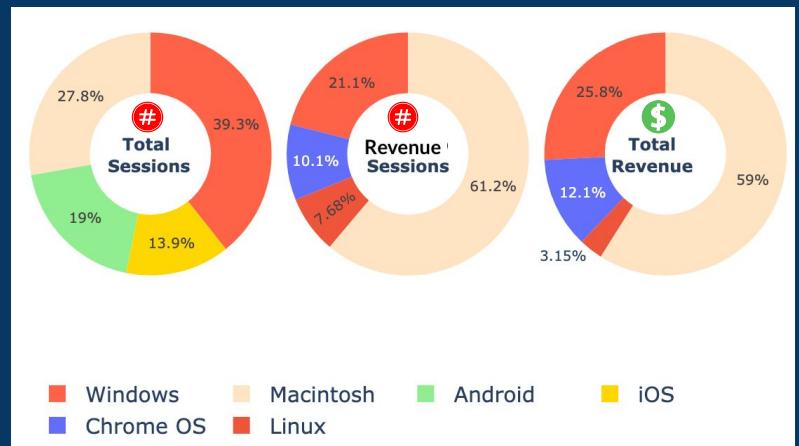
- ❖ Visits: shows the number of times a user visited store.
- ❖ Hits: shows the number of times a store (or) product page URL has accessed.

3. Most of the Sessions are through mobile devices

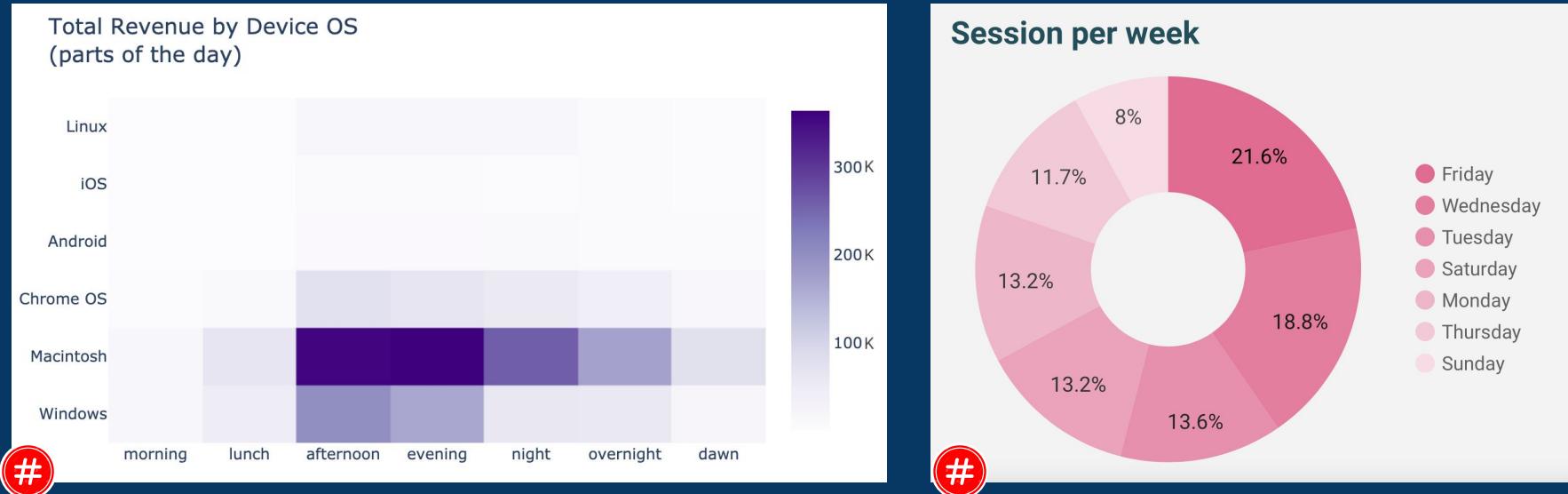
Total number of Sessions per channel:



Mac Users Sessions are higher in number and generate more Revenue

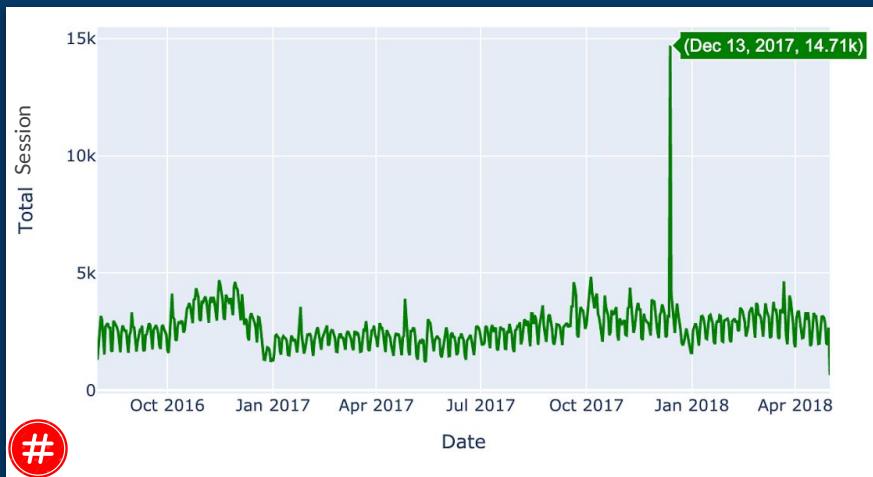


No. of Revenue Session at different parts of Day and Week

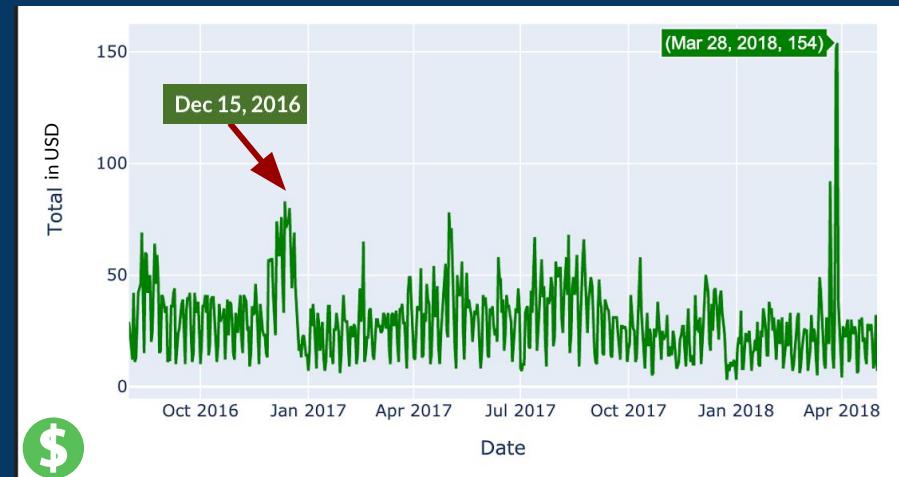


No. of Sessions per Date

Total Sessions over time (Rev + Non Rev):



Revenue based Session over time:



- High Sessions recorded in December 2017
- High peek on March 2018 for Sessions with Revenue!

Highlights Insights

- The **USA market** is the most profitable.
- **Organic search** generate the most traffic.
- Most number of the Revenue Transactions come from **Referral**.
- **Mac** -users are the strongest in terms of revenue, regardless the channel.
- Higher session traffic on **Fridays** and in **Afternoon and Evenings**.

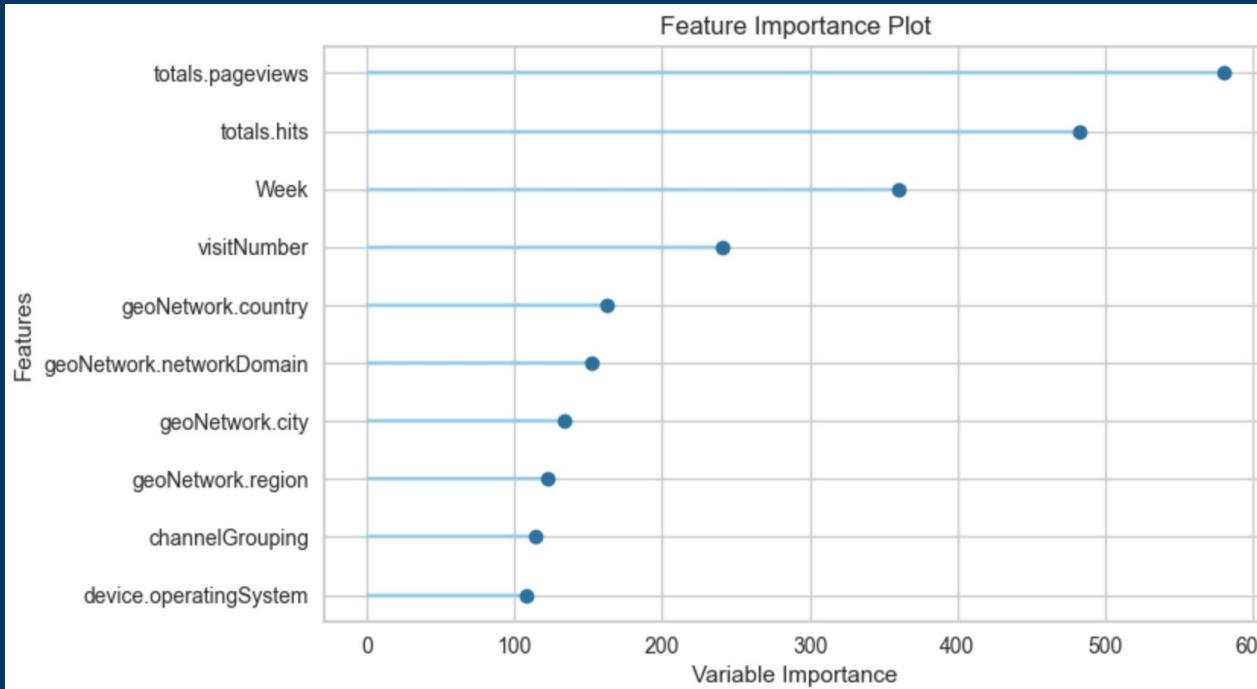
Agenda:

- ❖ Business Case
- ❖ Objectives & Approach
- ❖ Our Workflow
- ❖ Data Description
- ❖ Metric
- ❖ Exploratory Data Analysis
- ❖ Modeling
- ❖ Future Work

Methodology and Modeling

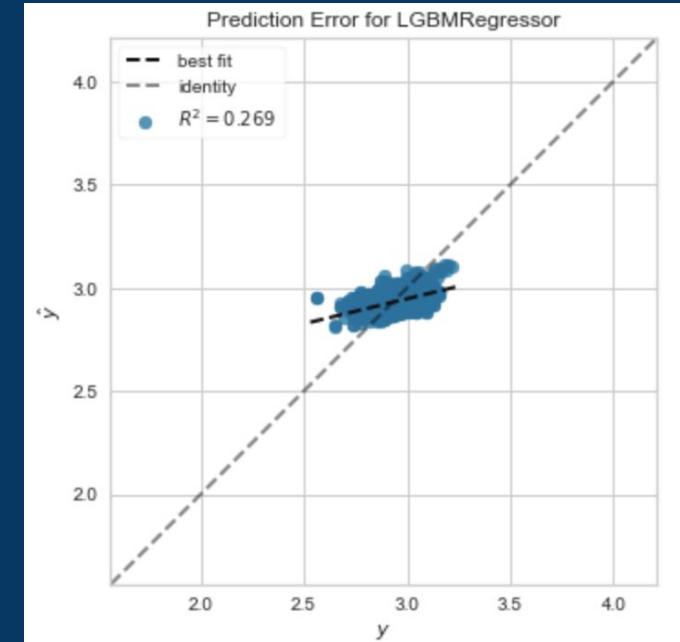
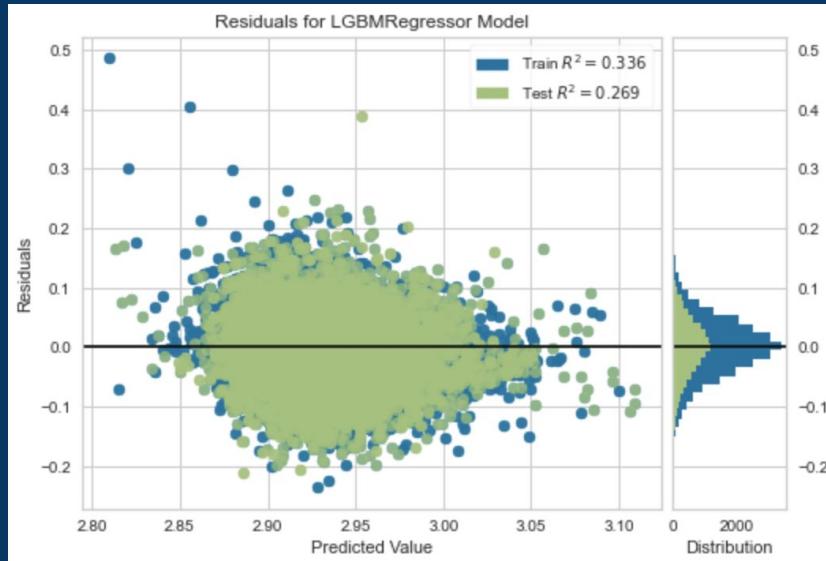
- Classification Model (Revenue or non-Revenue Sessions).
- Regression Model (The Value of Sessions Revenue).
- Calculating Target Value by multiplying results from two models.
- Aggregate predicted Total Revenue Transactions per Customers.

Feature Importance for our Model

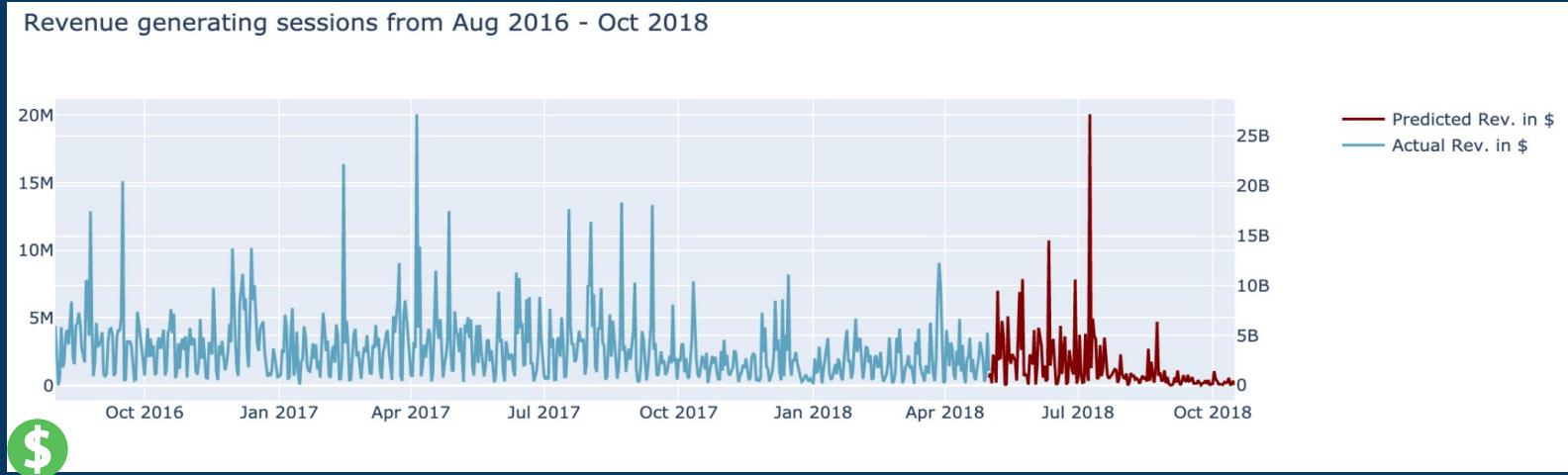


Error Analysis for our final model

Revenue generating analysis

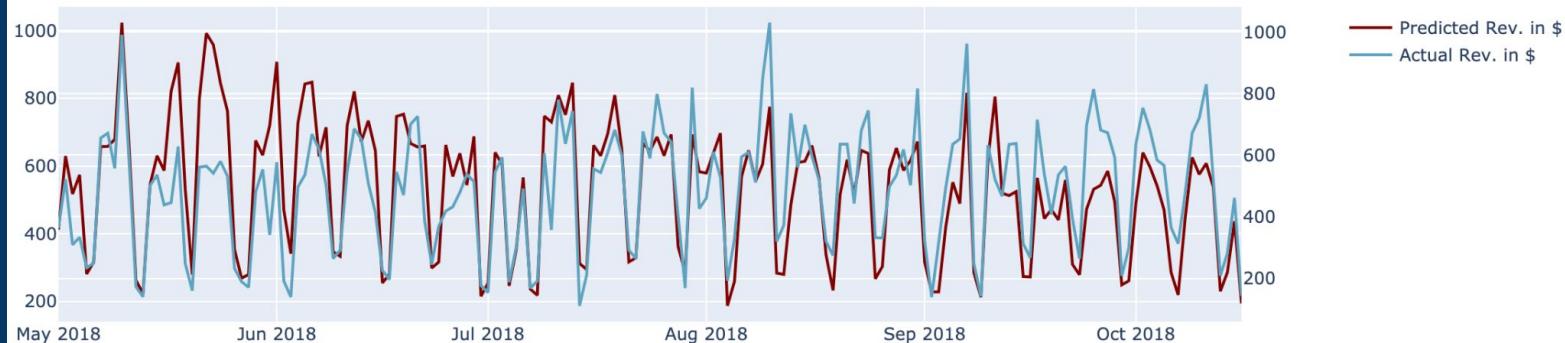


Results of the Model



How well did our model predict?

Result: Revenue generating sessions from May - Oct 2018



Our model RMSE: 1.71

Improved: 16.5%

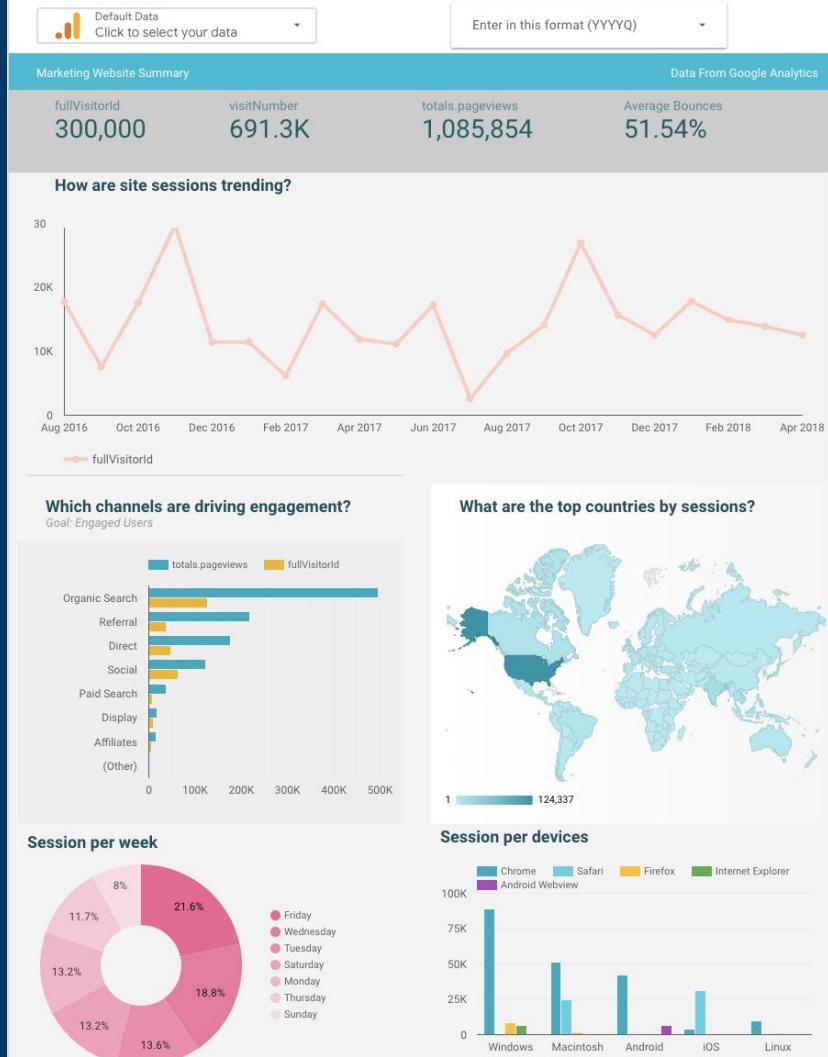
Baseline RMSE: 2.05

Agenda:

- ❖ Business Case
- ❖ Objectives & Approach
- ❖ Our Workflow
- ❖ Data Description
- ❖ Metric
- ❖ Exploratory Data Analysis
- ❖ Modeling
- ❖ Future Work

Next steps

- Our interactive dashboard is available on Google data-studio for further exploration
- Deploy the model to predict if the session brings Revenue



Questions & Feedback

Hassan Salam

- ❖ [Linkedin](#)
- ❖ [GitHub](#)



Jonathan Preiss

- ❖ [Linkedin](#)
- ❖ [GitHub](#)



Thank you for your attention!! :)

Sina Lotfimran

- ❖ [Linkedin](#)
- ❖ [GitHub](#)



Davain Pablo Edwards

- ❖ [Linkedin](#)
- ❖ [GitHub](#)

