



URANUS

Online Shoppers Purchasing Intention Dataset

WHO ARE WE?



Uranus

is an analyst team, who is working under Data Science department in the e-commerce company named Kaospedia

Kaospedia

is an independent ecommerce company focused on retail fashion product.

MEET WITH OUR TEAM



Rijal



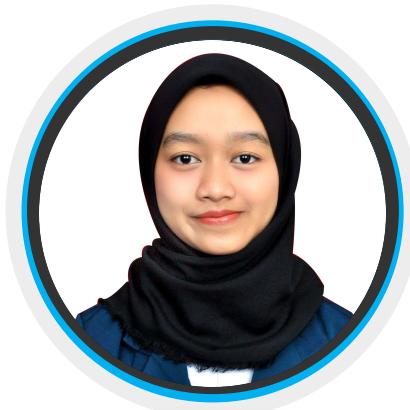
Zamzam



Sahel



Yusuf



Putrini



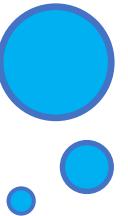
Rendra



Surya

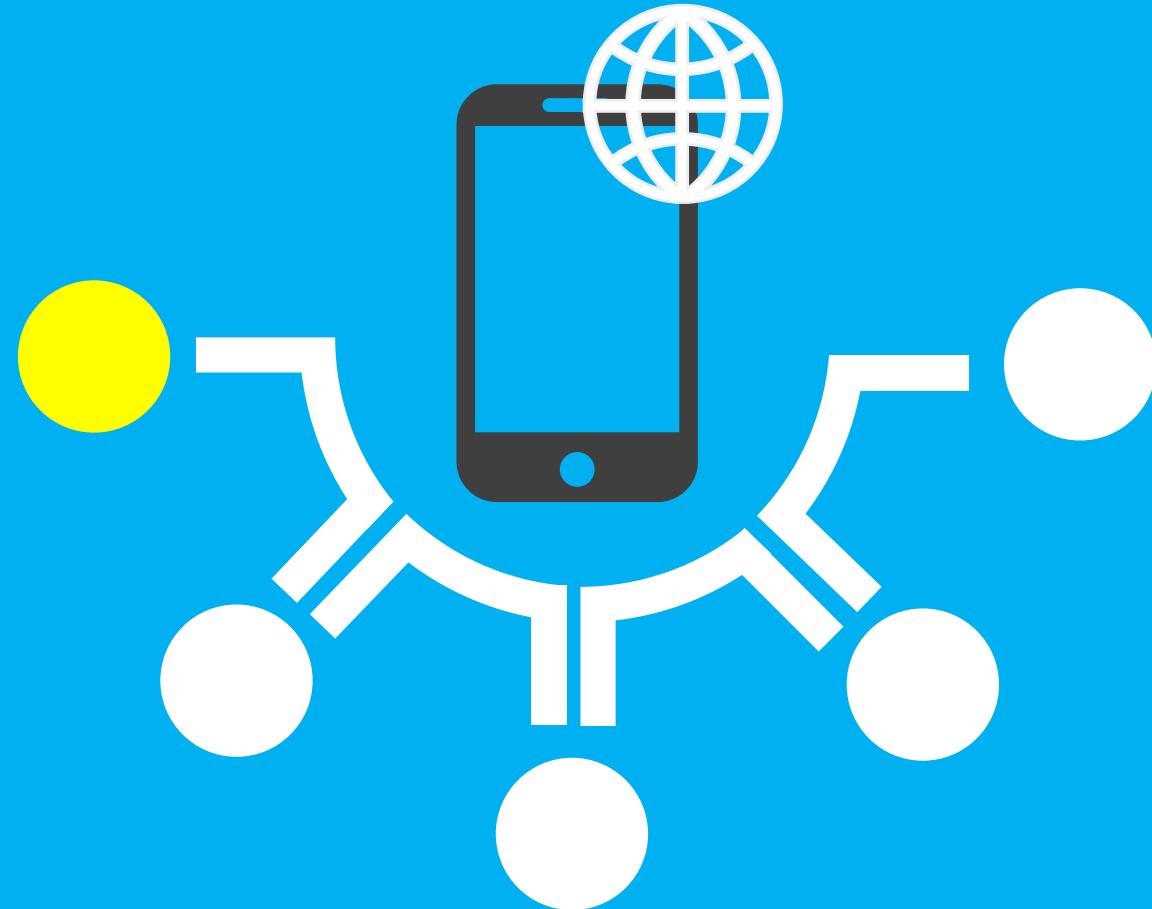
Table Of Content

- 01** **Problem Statement**
We need to analyze why out of the many visitors on this online shoppers website only 15% make a purchase. This result is relatively small.
- 02** **Goal, Objective, and Business Metrics**
We will increase intention rate from 15% to 25% by creating a machine learning model that will predict the criteria for visitors to this website. The metric we will be using is 'Revenue'
- 03** **EDA and Pre Processing**
Before build a model machine learning, we will handling missing values, duplicate data, and feature engineering.
- 04** **Building Machine Learning (Modelling)**
We will build some machine learning according to the data, and will choose the best version to apply for new data.
- 05** **Business Recommendation**
We will provide business recommendations that can be applied to have a real impact on this case.

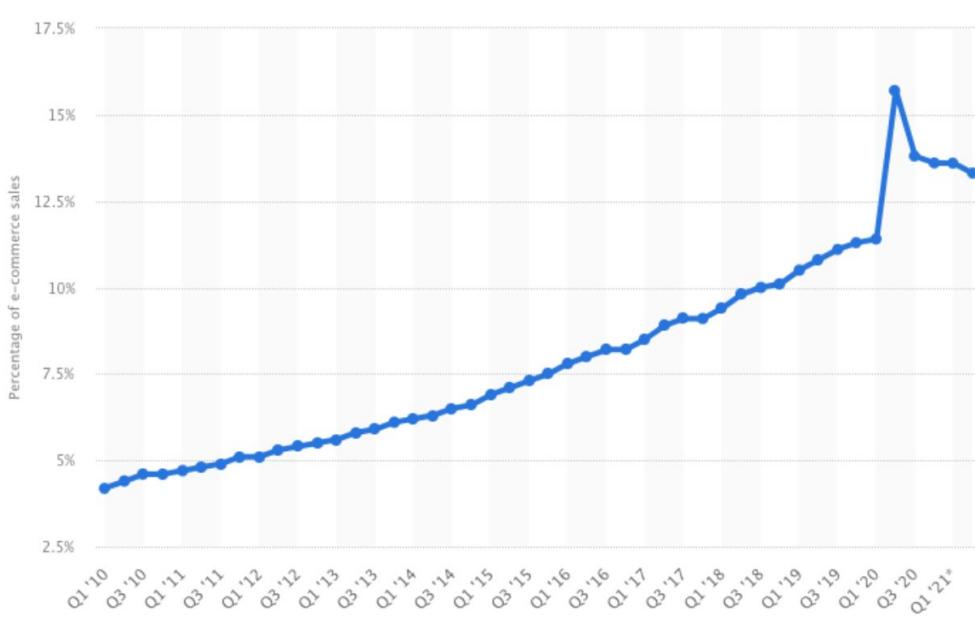


Problem Statement

Problem Statement



Research said



Percentage of e-commerce retail sales in the UK rising to over a third of all retail in the UK in November 2020 and January 2021.

It has since decreased slightly, but it will be interesting to follow this trend in the run-up to the 2022 holiday season, and into next year.

Problem Statement



- The number of online shoppers worldwide are always increasing each year.
- This indicate that the competition between e commerce company is getting tougher.

Problem Statement

The conversion rate is the number of visitors that complete a transaction divided by the total number of visitors represented in a percentage

$$\text{Conversion rate} = \frac{\text{Number of transaction}}{\text{Number of website visit}} \times 100\%$$

Based on studies, average ecommerce conversion rate is between 20% and 40%.

Based on EDA, the current conversion rate in Kaospedia is

15.47%



- Therefore, company management set the target higher to 25% to increase company competitiveness.

Conversion rate target

25 %

Problem Statement



Current Condition

Conversion Rate = 15.47%

We don't want sales to decline next year, we will give solution with analyzing and providing the best business recommendations to improve revenue



Target



What will we do?

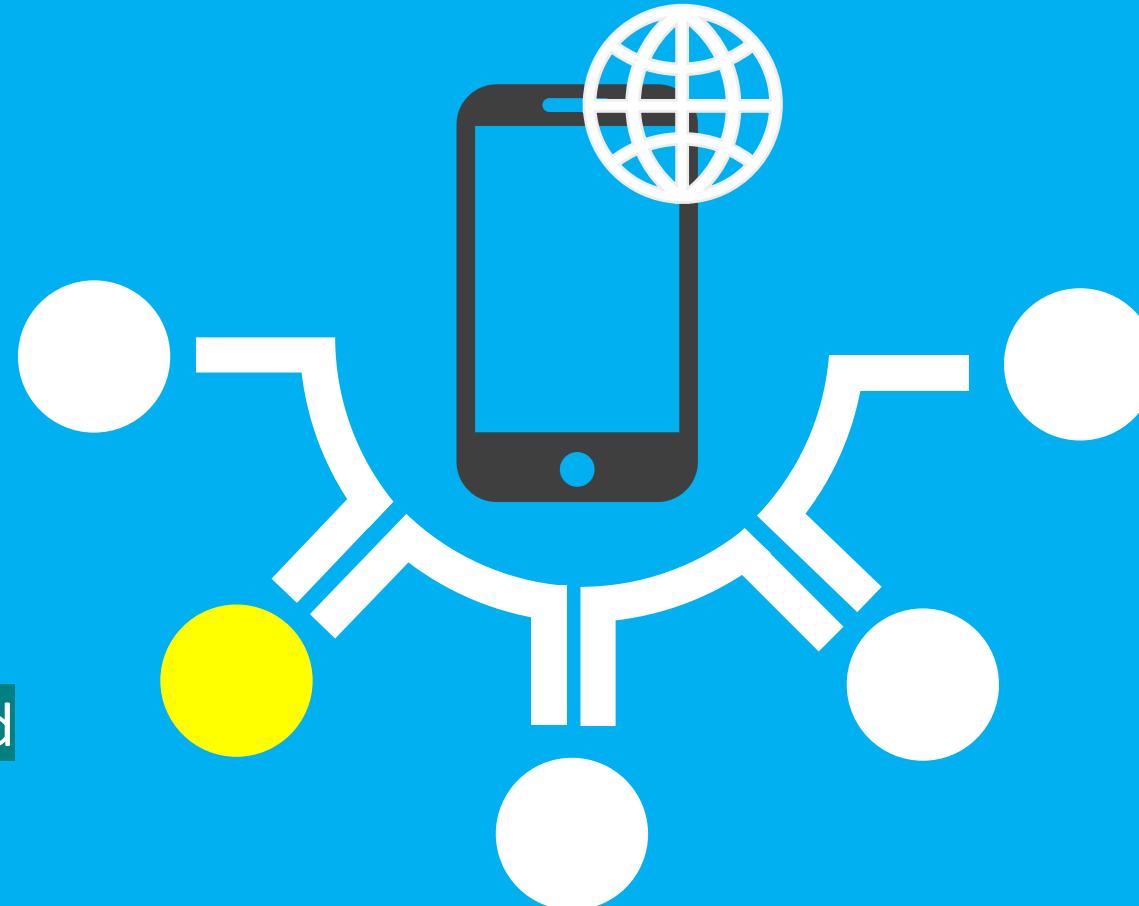
Conversion rate = 25%

- ✓ analyze data and get the actionable insights
- ✓ create business recommendation using both descriptive and inferential
- ✓ automate the visitor behavior prediction by machine learning model

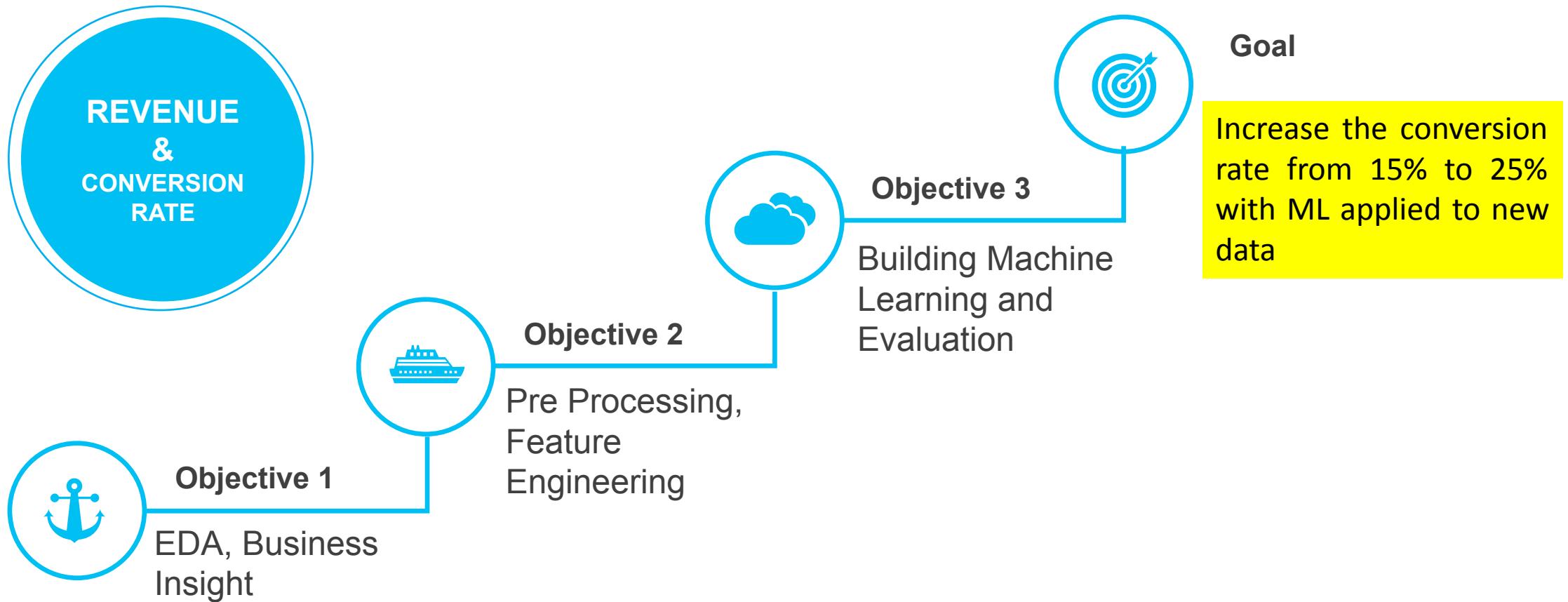
Goal, Objective, and Business Metrics

Problem Statement

Goal, Objective, and
Business Metrics



Business Metrics

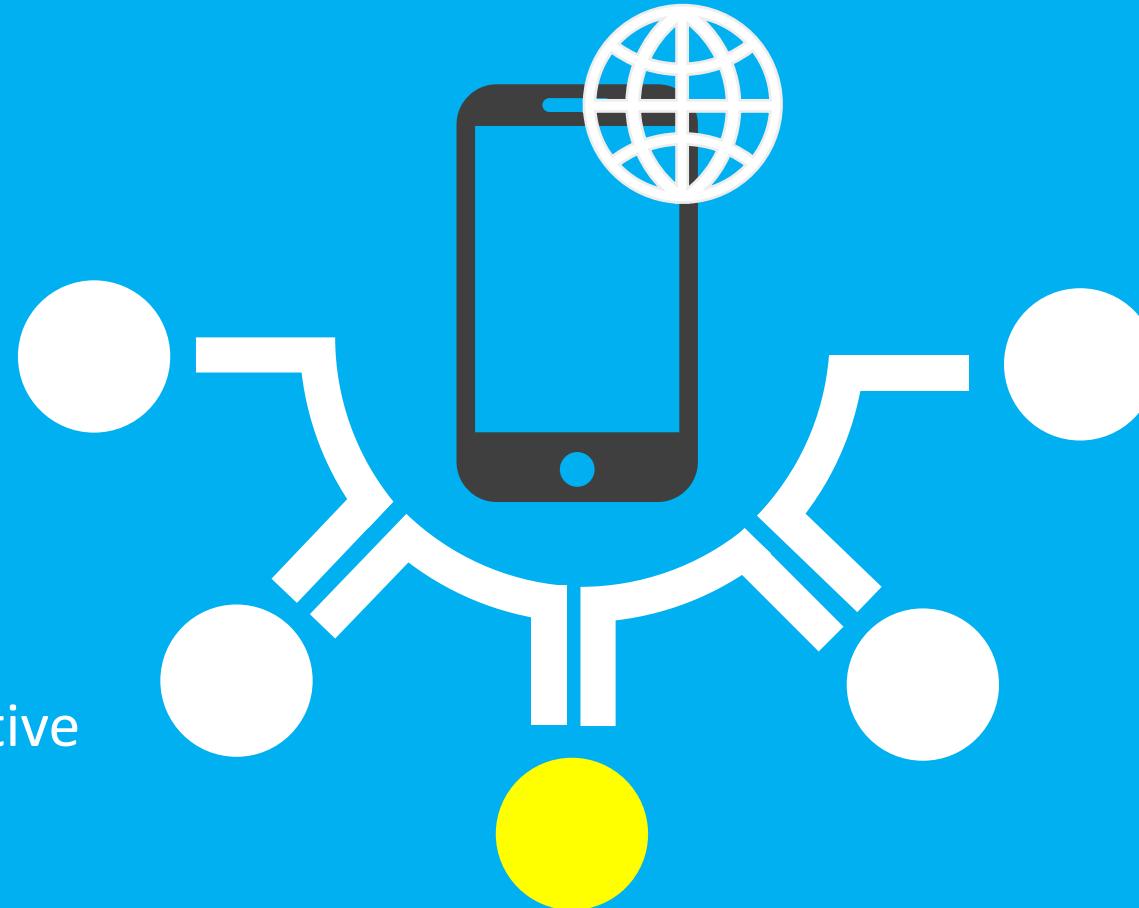


Pre Processing and EDA

Problem Statement

Goal and Objective

Pre Processing and EDA

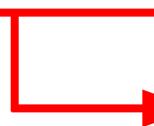


DATASET OVERVIEW

No	Column	Descriptive	Type Data	Missing Value	Remark	
1	Administrative	This is the number of pages of this type (administrative) that the user visited.	Numerical	0	min : 0 max : 27	median : 1 mean : 2.315166
2	Administrative Duration	This is the amount of time spent in this category of pages.	Numerical	0	min : 0 max : 3398.75	median : 7.5 mean : 80.818611
3	Informational	This is the number of pages of this type (informational) that the user visited.	Numerical	0	min : 0 max : 24	median : 0 mean : 0.503569
4	Informational Duration	This is the amount of time spent in this category of pages.	Numerical	0	min : 0 max : 2549.375	median : 0 mean : 34.472398
5	Product Related	This is the number of pages of this type (product related) that the user visited.	Numerical	0	min : 0 max : 705	median : 18 mean : 31.731468
6	Product Related Duration	This is the amount of time spent in this category of pages.	Numerical	0	min : 0 max : 63973.5222	median : 598.936905 mean : 1194.74622
7	BounceRates	The percentage of visitors who enter the website through that page and exit without triggering any additional tasks.	Numerical	0	min : 0 max : 0.2	median : 0.003112 mean : 0.022191
8	ExitRates	The percentage of pageviews on the website that end at that specific page.	Numerical	0	min : 0 max : 0.2	median : 0.025156 mean : 0.043073
9	PageValues	The average value of the page averaged over the value of the target page and/or the completion of an eCommerce	Numerical	0	min : 0 max : 361.763742	median : 0 mean : 5.889258

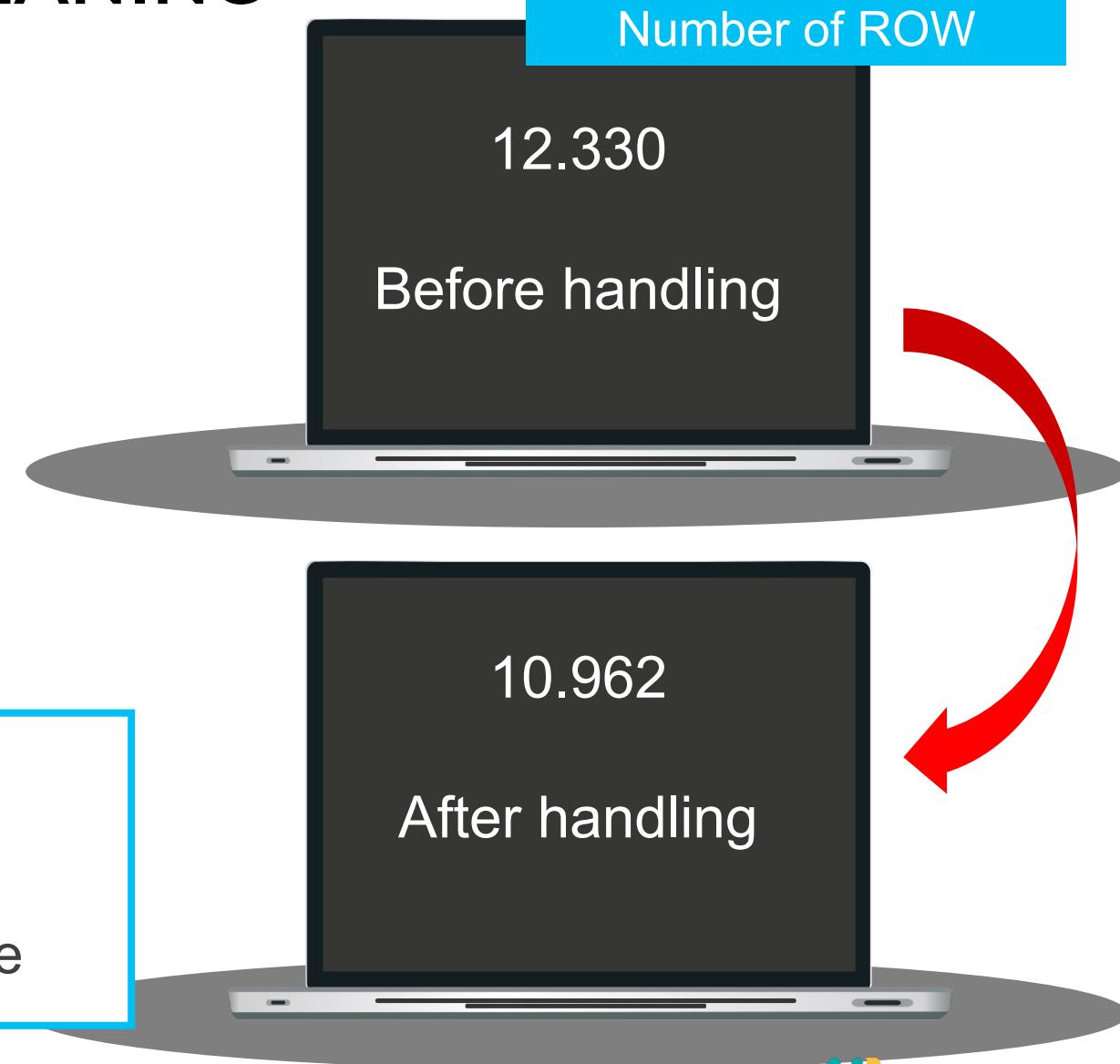
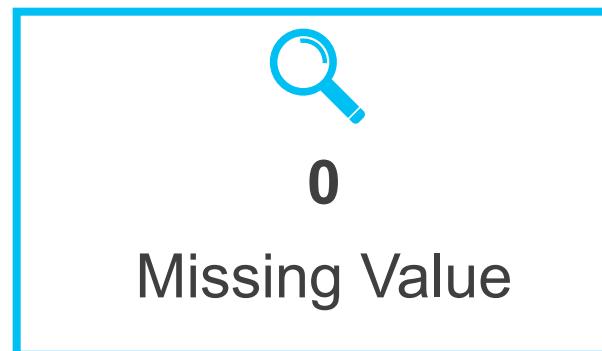
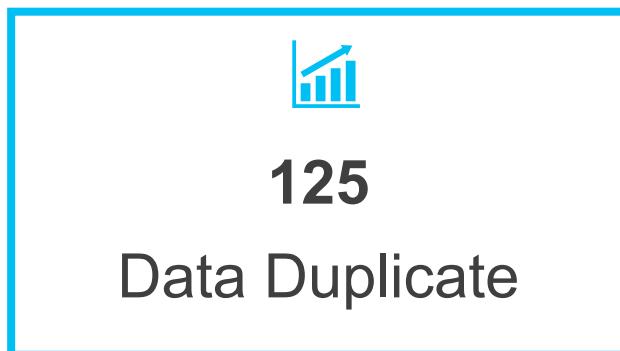
DATASET OVERVIEW

No	Column	Descriptive	Type Data	Missing Value	Remark	
10	SpecialDay	This value represents the closeness of the browsing date to special days or holidays (eg Mother's Day or Valentine's day)	Numerical	0	min : 0 max : 1	median : 0 mean : 0.061427
11	Month	Month value of visit date	Categorical	0	10 unique values	
12	OperatingSystems	Operating System of the visitor	Categorical	0	8 unique values	
13	Browser	Browser visitor use to open website	Categorical	0	13 unique values	
14	Region	RGeographic region from which the session has been started by the visitor	Categorical	0	9 unique values	
15	TrafficType	Traffict source by visitor	Categorical	0	20 unique values	
16	VisitorType	Kind of visitor type (New visitor / R)	Categorical	0	3 unique values - Returning_visitor - New_Visitor - Other	
17	Weekend	Weekend or Weekdays	Categorical	0	2 unique values : [true, false]	
18	Revenue	Visitor purchasing or not	Categorical	0	2 unique value : [true, false]	

 **TARGET**

DATA CLEANING

- Handling Duplicates
- Handling Outlier
- Handling 4 Column Redundant



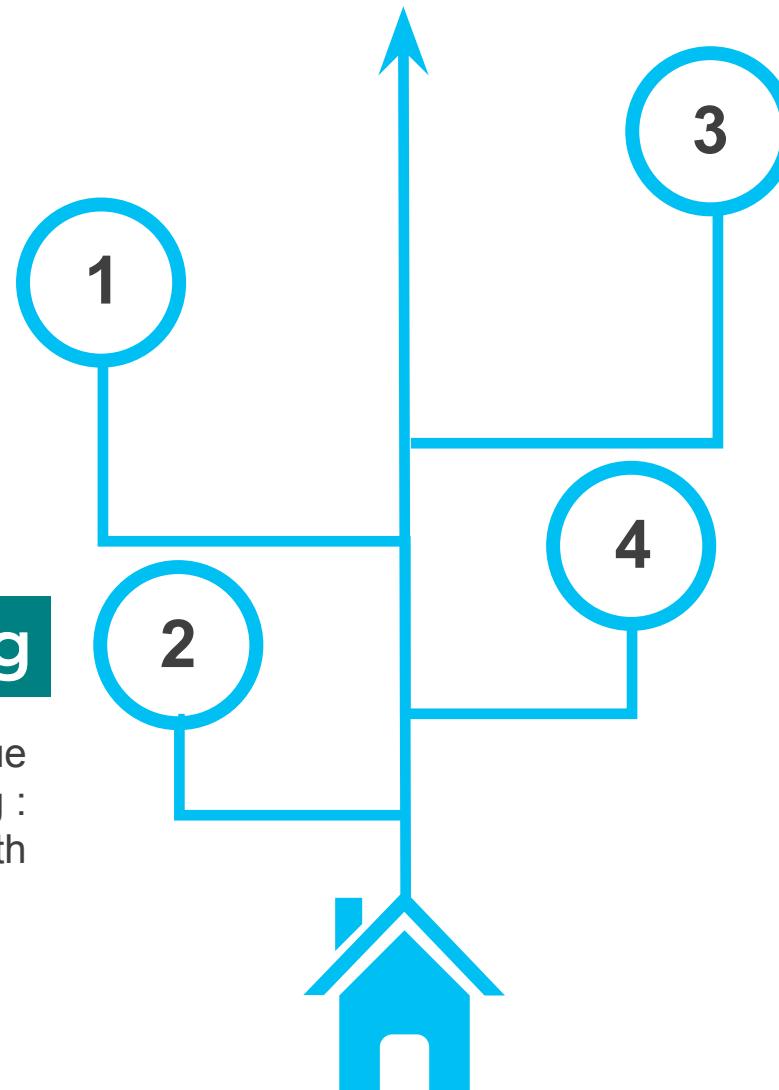
DATA PREPROCESSING

Feature Transformation

Normalisasi : Exit Rates
Standarisasi : administrative duration, informational duration, product related duration, page values, special day, traffic type.

Feature Encoding

Label Encoding : Revenue
One Hot Encoding : Visitor Type, Month



Feature Selection

Drop column :
Administration, ProductRelated,
Informational, BounceRates

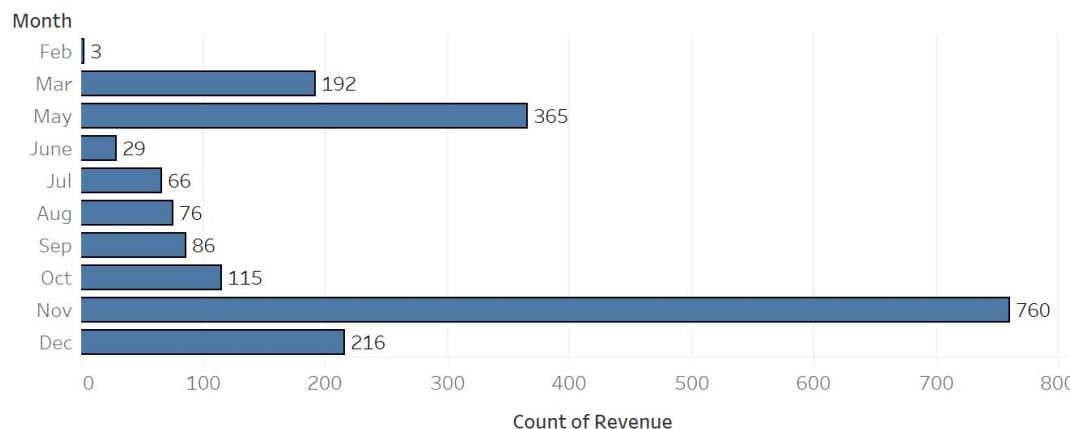
Feature Engineering

Oversampling SMOTE



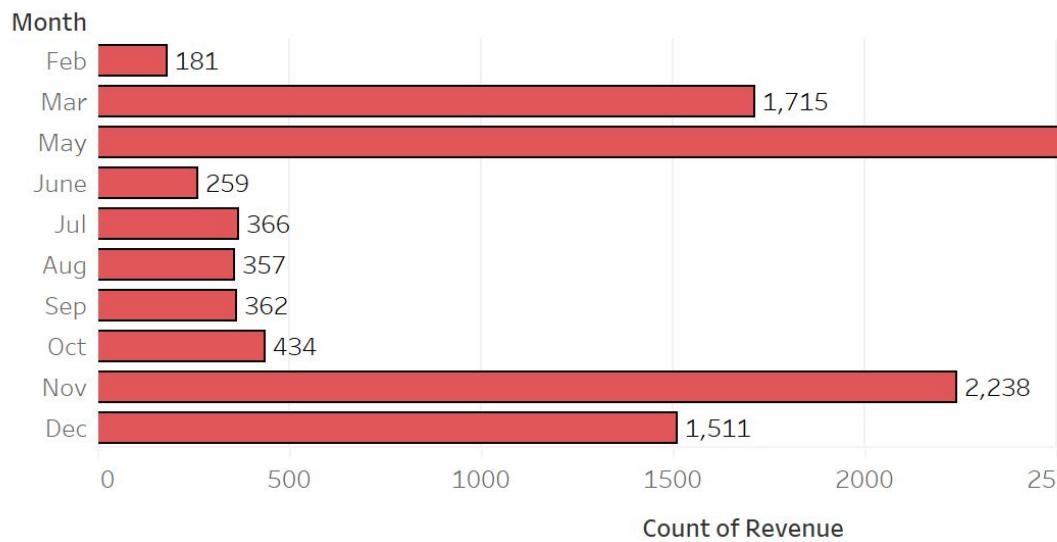
EDA

True Revenue



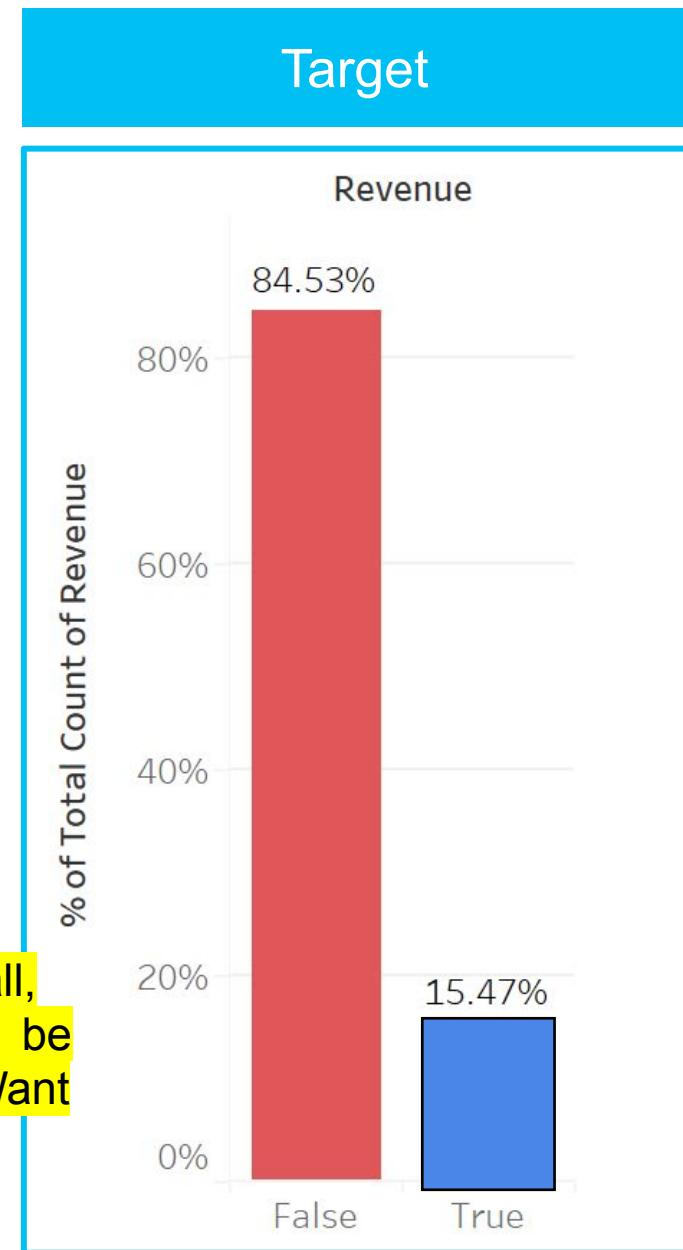
November and May have the most visitors and highest revenue

False Revenue

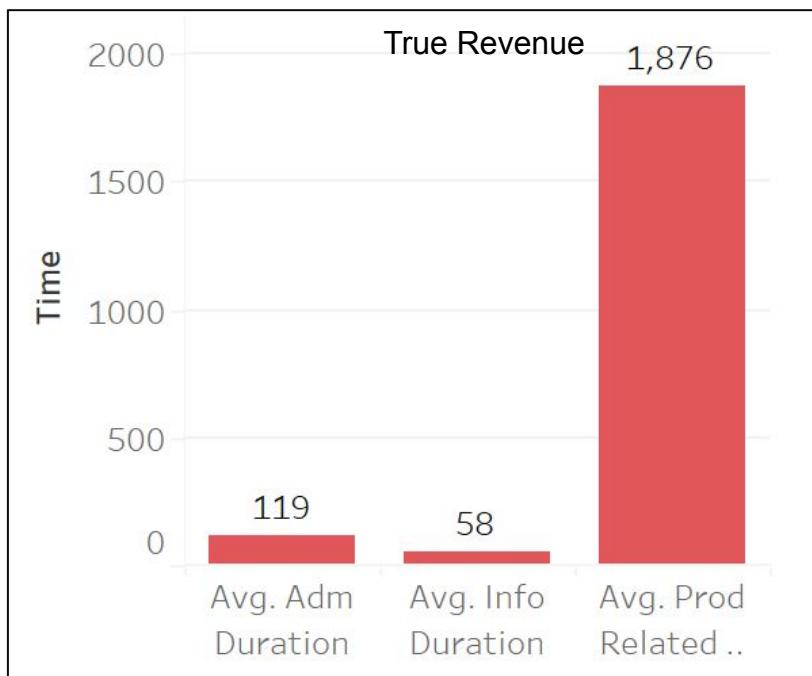


15% is relatively small, And still can be improved. We Want growth

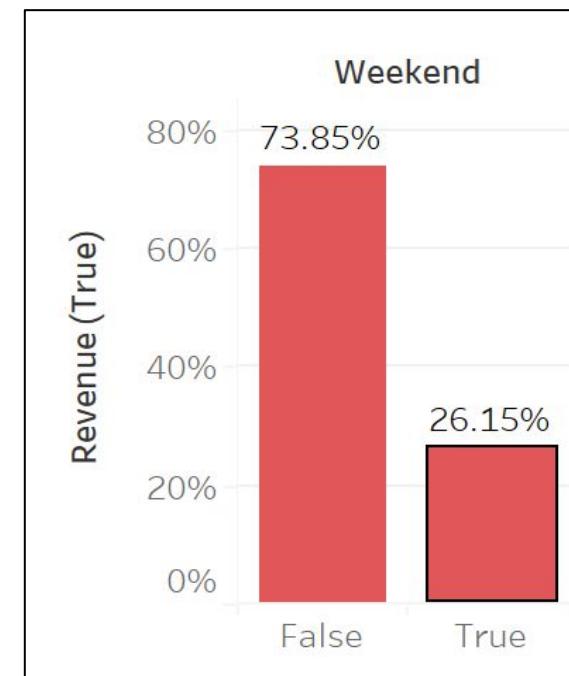
Target



EDA

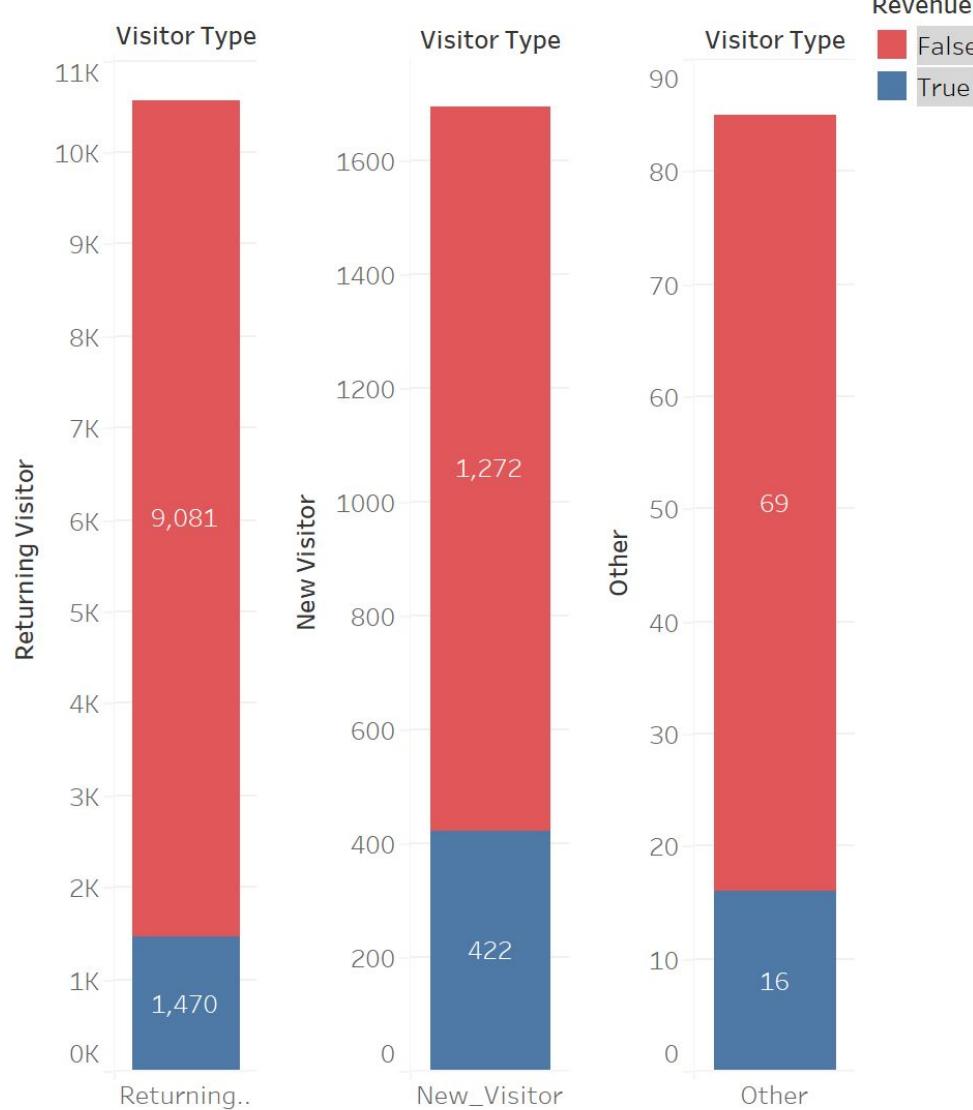


Visitor takes 30 minutes on related product before converting revenue

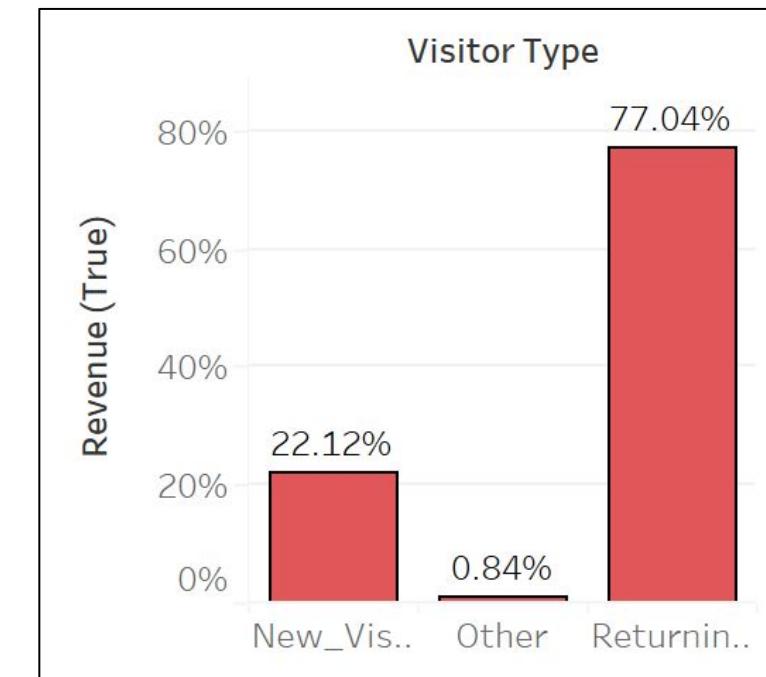


Visitor on Weekdays convert revenue almost 3 times more compared to Visitor on Weekend

EDA



Percentage new visitor have a higher than returning visitor based on revenue(true or false)



Returning Visitor convert revenue 3.5 times more compared to New visitors.

EDA OUTCOME

DESCRIPTIVE BUSINESS INSIGHTS



- E-commerce websites are dominated by visitors who have visited the website before and may have made a purchase 3X times more
- The highest sales occurred in November and May
- The average visitor takes 30 minutes to choose a product before deciding to buy
- Visitors who make purchases through the website are 15% of the total visitors (annual)

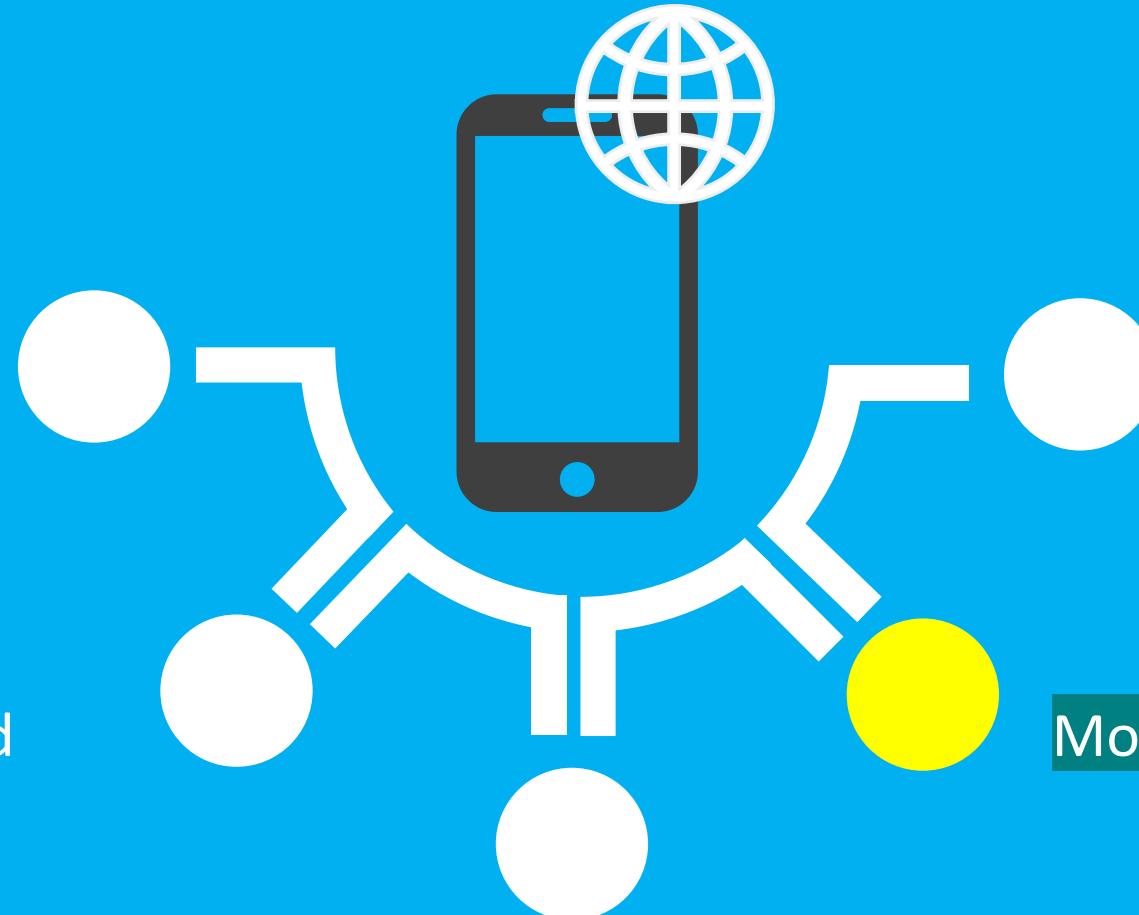
Modelling

Problem Statement

Goal, Objective, and
Business Metrics

Pre Processing and
EDA

Modelling



MODELLING

Model	AUC		Accuracy		Precision		TT (sec)
	Train	Test	Train	Test	Train	Test	
LightGBM	93.36%	88.87%	89.30%	80.53%	97.33%	94.53%	0.403871
CatBoost	94.32%	88.87%	90.93%	80.27%	97.76%	94.44%	6.981506
XGBoost	88.89%	88.00%	85.46%	82.23%	96.70%	95.33%	1.347314

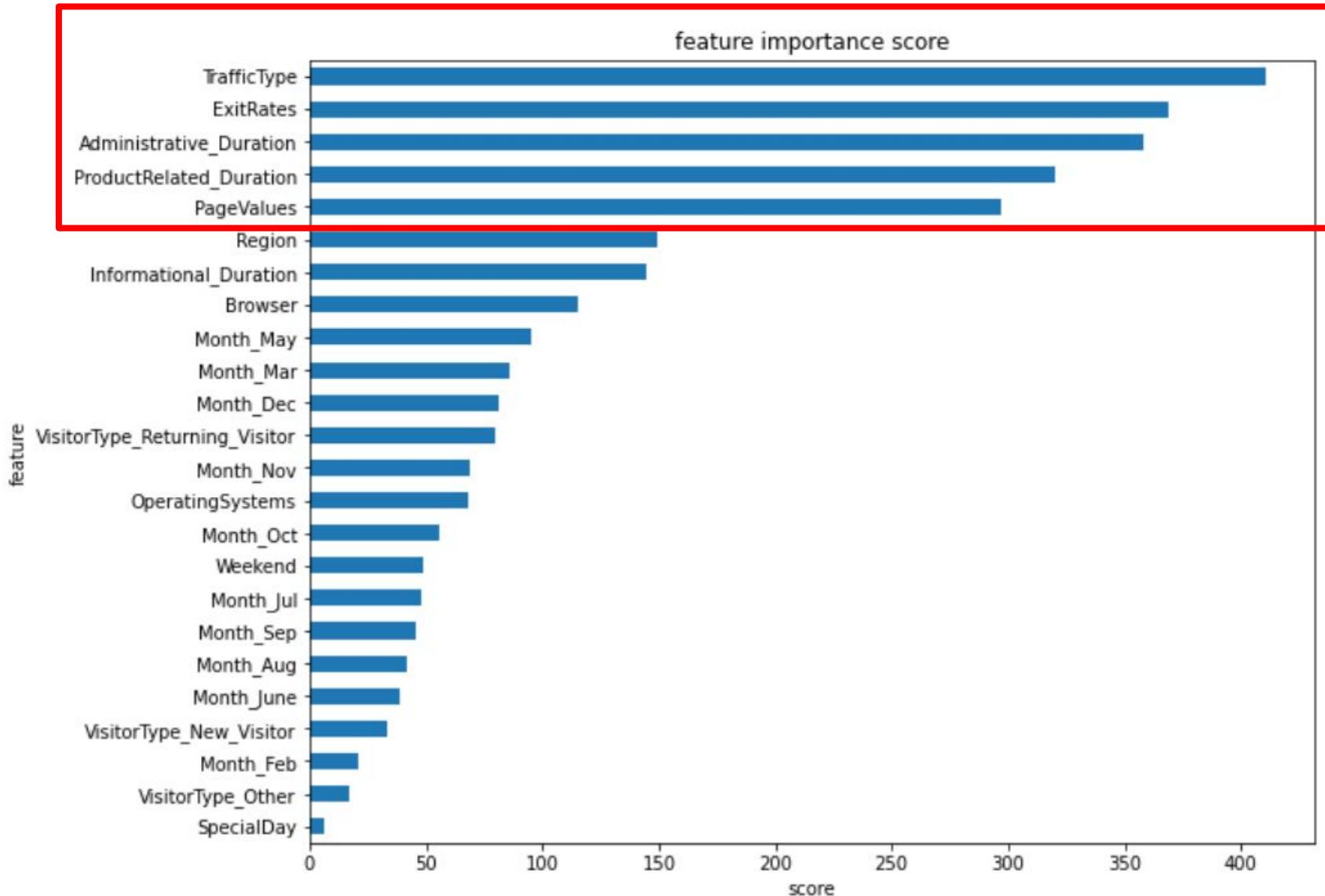
Our team focuses on metric **precision** for **Revenue False** values because our team wants to reduce **False Negatives** so that the number of purchases from visitors can increase

Dataset split ratio 80:20 (train:test)

Based on the results of the comparison of models by considering metrics, the **Light GBM** model is our choice because it has a **high precision value, high AUC value, and relatively low execution time.**

From the **AUC and Precision values** on the LightGBM model which are not significantly different, we conclude that this model is quite the **best-fit**

FEATURE IMPORTANCE

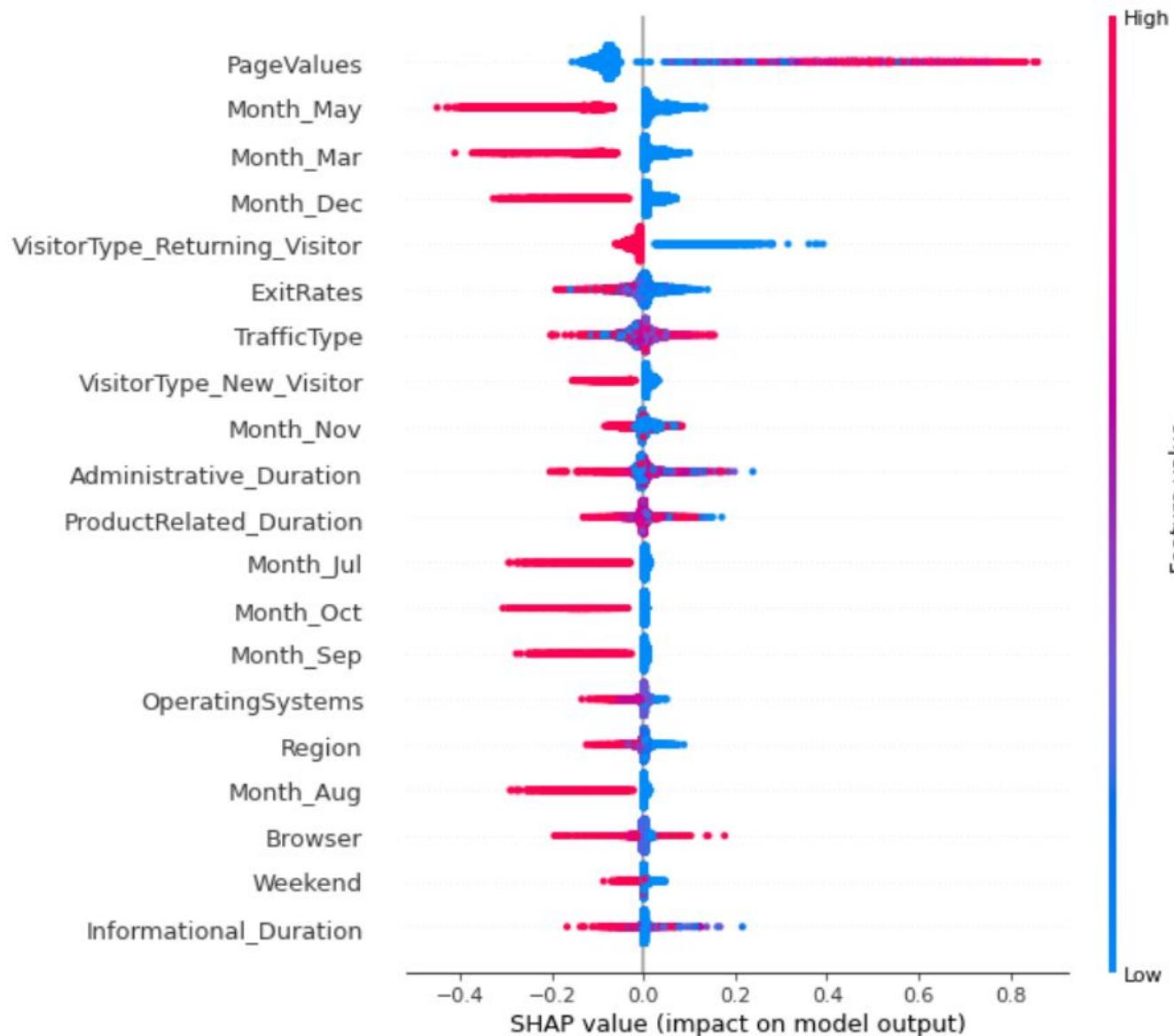


Top Five Feature Importance :

- Traffic Type
- Exit Rates
- Administrative Duration
- Product Related
- Page Value

Returning Visitor have 2 time more impacted than New Visitor

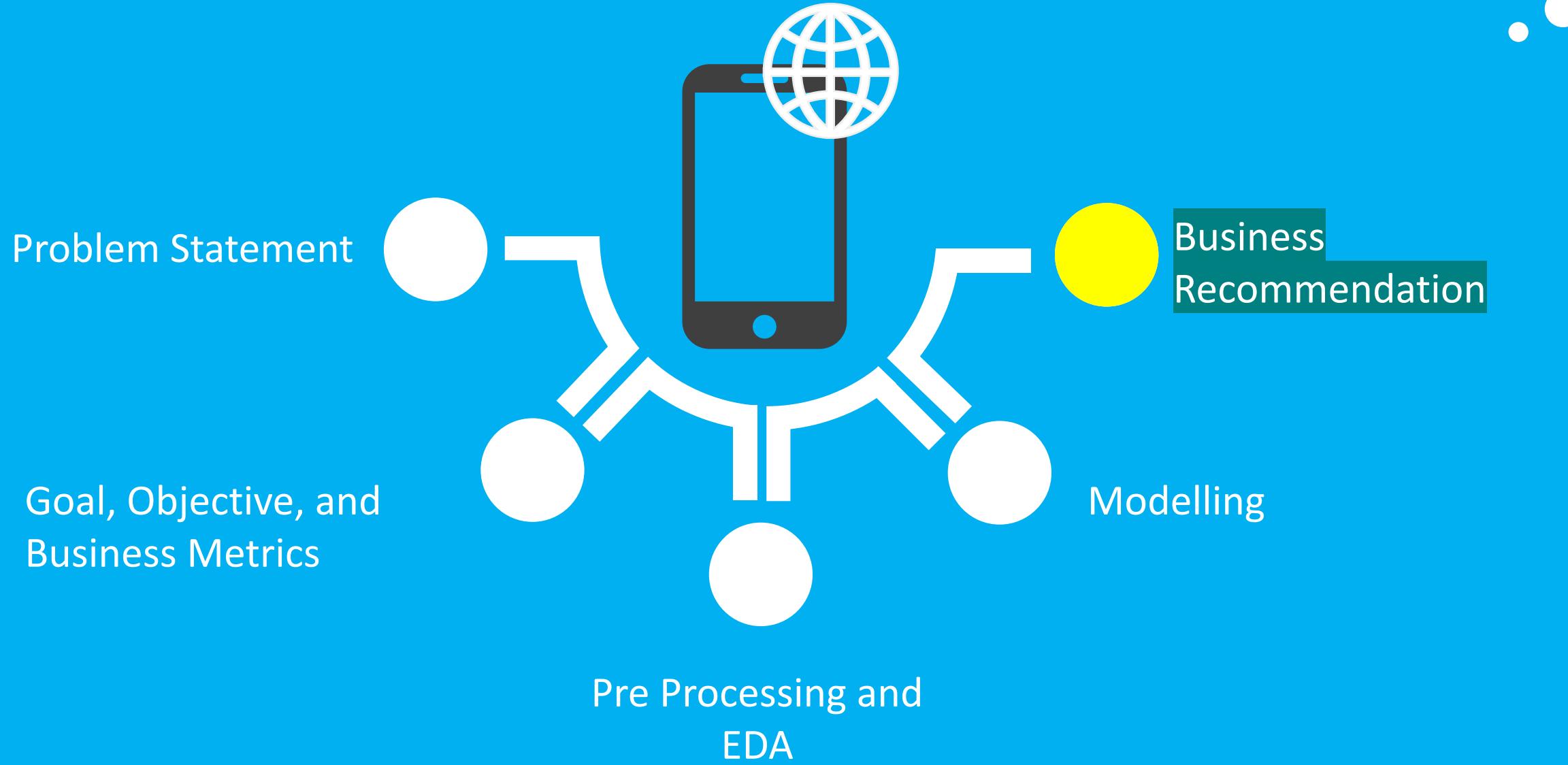
SHAP VALUE



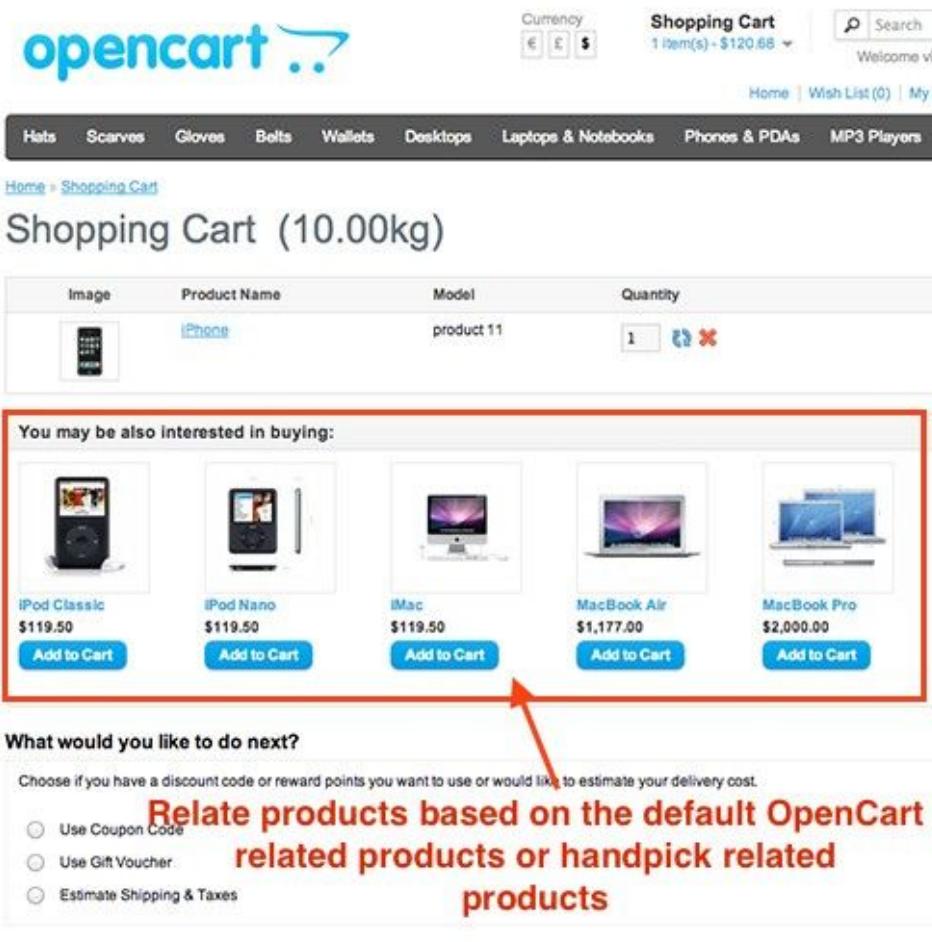
Users who open a lot of pages have a high potential to make a purchase compared to only a few

Type 20 traffic type (highest) has a higher purchase potential than low type (1,2..), based on research, traffic with the highest conversion rate is from: Social Media

Business Recommendation



BUSINESS INSIGHT AND RECOMMENDATION



The screenshot shows a shopping cart page from an OpenCart e-commerce platform. At the top, there's a navigation bar with links for Hats, Scarves, Gloves, Belts, Wallets, Desktops, Laptops & Notebooks, Phones & PDAs, and MP3 Players. The main content area displays a single item in the cart: a 'Phone' (product 11) with a quantity of 1. Below this, a section titled 'You may be also interested in buying:' contains five recommended products: iPod Classic (\$119.50), iPod Nano (\$119.50), iMac (\$119.50), MacBook Air (\$1,177.00), and MacBook Pro (\$2,000.00). Each recommendation includes an 'Add to Cart' button. A red box highlights this recommendation section, and a red arrow points from the text 'Relate products based on the default OpenCart related products or handpick related products' to the 'Add to Cart' button for the iMac.

opencart ..

Currency: € ₩ \$

Shopping Cart: 1 Item(s) - \$120.68

Search

Welcome visitor

Hats Scarves Gloves Belts Wallets Desktops Laptops & Notebooks Phones & PDAs MP3 Players

Home > Shopping Cart

Shopping Cart (10.00kg)

Image	Product Name	Model	Quantity
	Phone	product 11	1

You may be also interested in buying:

	iPod Classic \$119.50	
	iPod Nano \$119.50	
	iMac \$119.50	
	MacBook Air \$1,177.00	
	MacBook Pro \$2,000.00	

What would you like to do next?

Choose if you have a discount code or reward points you want to use or would like to estimate your delivery cost.

Use Coupon Code
 Use Gift Voucher
 Estimate Shipping & Taxes

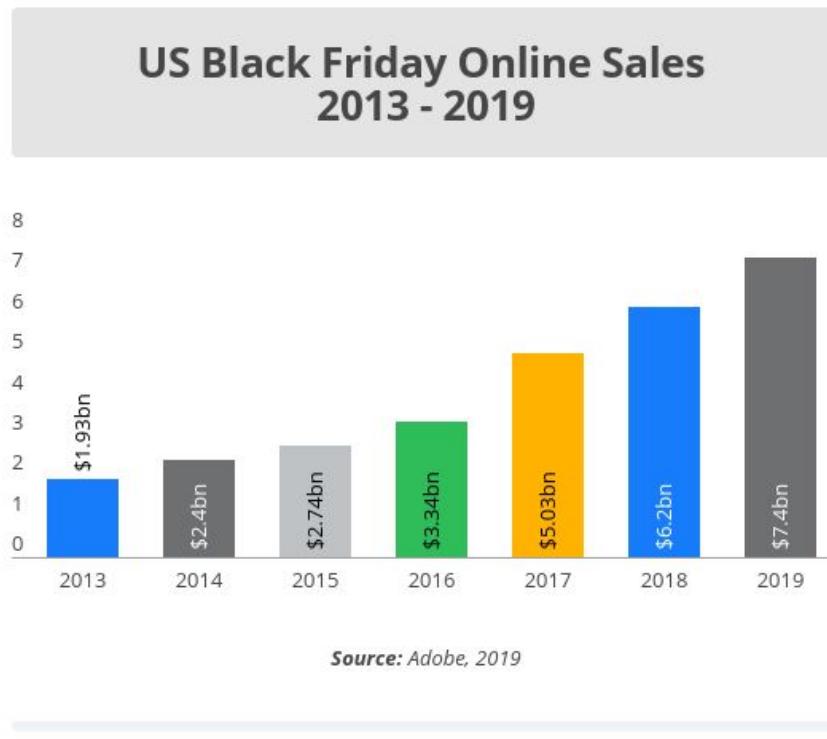
Relate products based on the default OpenCart related products or handpick related products

An average visitor opens 6 pages before making a purchase



Adding 5 product recommendations relate to each product, so that customers linger on the website for a long time.

BUSINESS INSIGHT AND RECOMMENDATION



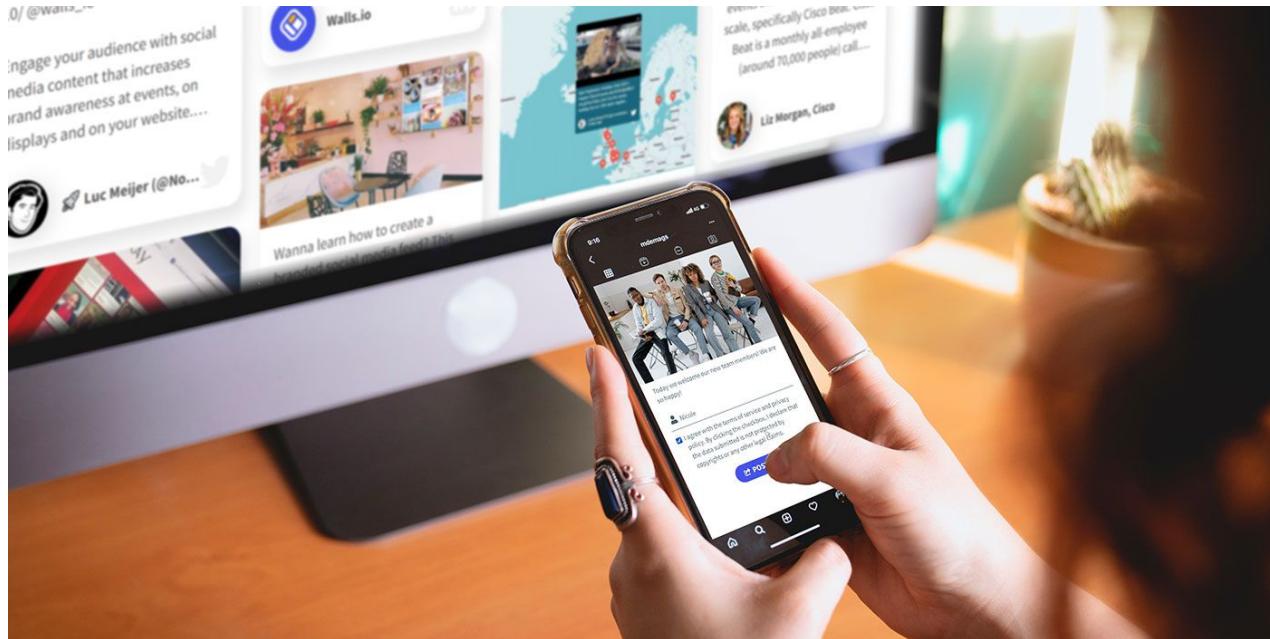
November sales
high

Advise the business team to bring high traffic to the website because there is Black Friday.



With this type of traffic from social media in the hope of taking advantage of the momentum to be able to increase sales

BUSINESS INSIGHT AND RECOMMENDATION



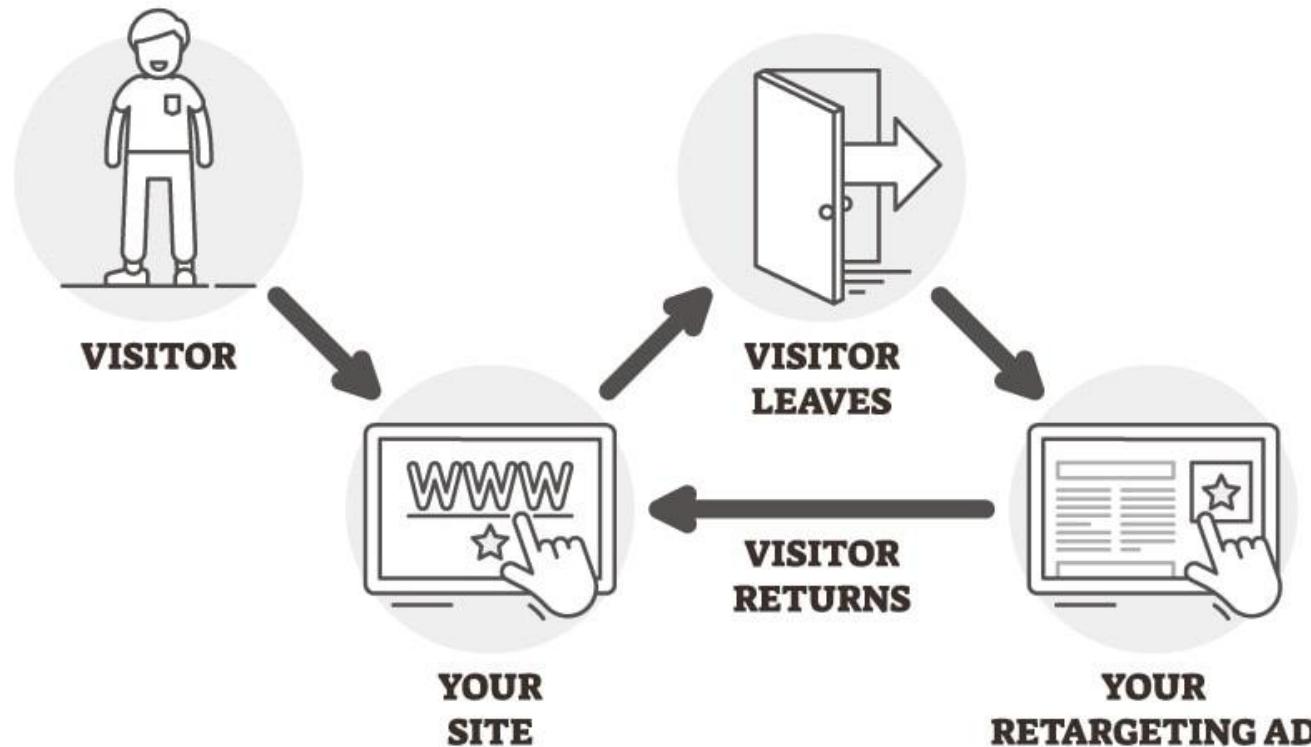
Social Media is the best traffic to generate conversion rate



Advise the business team to focus on increasing traffic from social media directly to the website

BUSINESS INSIGHT AND RECOMMENDATION

RETARGETING



Returning visitors are more likely to make purchases

Advise the business team to retarget users who have visited the website, but have not or have not made a purchase

SIMULATION

Assumption

The average revenue for each purchase transaction is Rp. 100,000
Total e-commerce visitors are 2193 visitors

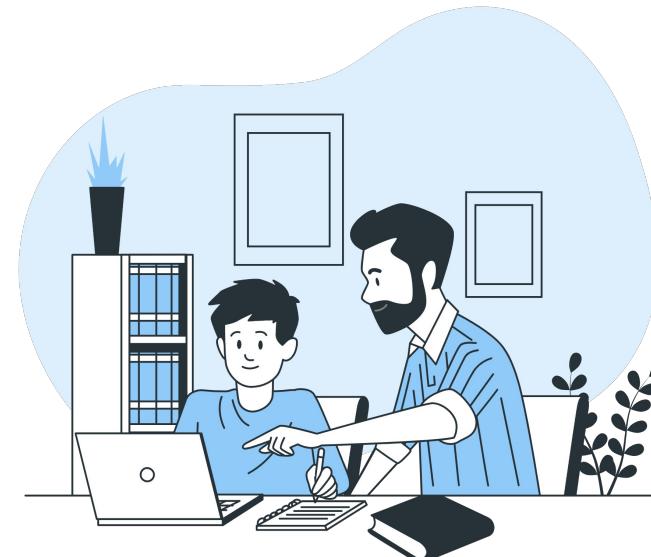
* This simulation uses data testing
** CVR in this case is the percentage of visitors who end up purchasing

BEFORE

The number of visitors who made purchases is
320 visitors

Total revenue = Rp 32.000.000 (320x Rp 100.000)

CVR
 41.22%
 14.59%



AFTER

TREATMENT BASED ON MODEL
The number of visitors who made purchases is
904 visitors

Total revenue = Rp 90.400.000 (904x Rp 100.000)

Total Revenue
 Rp 90.400.000
 Rp 32.000.000

ACTION ITEMS

In order to increase the conversion rate to 41.22% (as predicted by the ML model), we suggest the product and business team to apply the AARRR framework:

1. **Awareness and Acquisition:** Increasing traffic by bringing in the traffic from social media to e-commerce web (Traffic type)
2. **Activation:** Increasing traffic on November due to black Friday, maximize the marketing events on this month (Month)
3. **Retention:** Retargeting for returning visitors (Traffic Type, Visitor Type)
4. **Revenue:** Adding more product recommendation features on product related pages (Page values, Administrative Duration, Product related duration)
5. **Referral:** Make a referral campaign, get bonuses when the users are able to bring their friends to transact here (traffic type, page values)

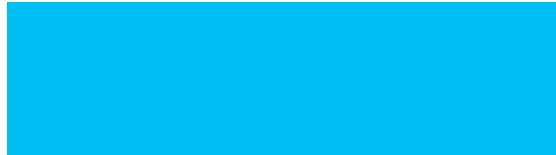




Thank you

BACK UP SLIDE

APPLICATION MACHINE LEARNING



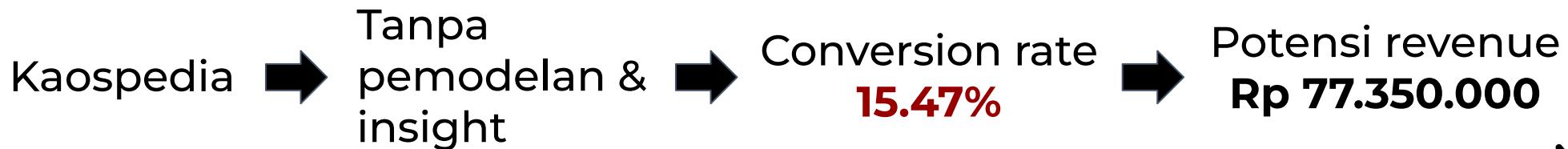






REVENUE CALCULATION

BEFORE



Asumsi

- Marketing cost Rp 10.000.000
- Cost perclick Rp 2.000
- Transaction amount percustomer Rp 100.000

AFTER



↑**61%**

MODELLING

Model	AUC		Accuracy		Precision		TT (sec)
	Train	Test	Train	Test	Train	Test	
LightGBM	0.933630	0.888737	0.893001	0.805309	0.973281	0.945325	0.403871
CatBoost	0.943209	0.888737	0.909337	0.802718	0.977552	0.944354	6.981506
XGBoost	0.888927	0.880073	0.854558	0.822261	0.966964	0.953266	1.347314

- Tim kami berfokus kepada metric **precision** untuk nilai **Revenue False** karena tim kami ingin **menurunkan False Negatif** agar angka pembelian dari pengunjung mampu meningkat
- Berdasarkan hasil perbandingan model dengan mepertimbangan metrics, model **Light GBM** menjadi pilihan kami karena memiliki nilai **Precision yang tinggi**, nilai **AUC yang tinggi**, dan **waktu eksekusi yang relatif rendah**.
- Dari nilai AUC dan Precision pada model LightGBM yang tidak berbeda secara signifikan, kami menyimpulkan bahwa model ini sudah **cukup best-fit**