



# Analysis of online shoppers intentions

What are the visitors's trends ?

Ulysse BOURDILLON  
Emmanuelle MEVANZA  
Hélène TE

DIA 4- ESILV

# Table of contents

01

## Introduction

Information about the dataset  
Problem statement

03

## Data pre-processing and modeling

Clustering, Regression models

02

## Data visualization

First step into the data

04

## Conclusion

Retrospection about the project





01

# Introduction

Presentation of the dataset and problems identified



# Online Shoppers Purchasing Intention



**12 330**  
sessions



**1**  
year



**16.5%**  
Session that ends  
with a transaction



**12**  
Variables with no  
missing values



## Additional information

### Nature of variables :

- Type of pages: *Administrative, Informational, ProductRelated*
- Time spent on these pages:  
*Administrative\_Duration, Informational\_Duration, ProductRelated\_Duration*
- Metrics measured: *BouncesRates, ExitRates, PageValues*
- Information about the user: *OperatingSystems, Browser, Region, TrafficType, VisitorType*
- Information about the period of the visit:  
*Weekend, Month, SpecialDay*
- Transaction: *Revenue*

### Note:

Missing information values (OS, Browser, Region, TrafficType)





**What are the visitors's trends ?**





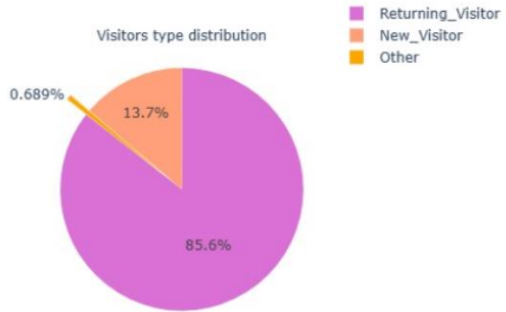
02

# Data visualization

You can enter a subtitle here if you need it



# First sight on the shopper's types

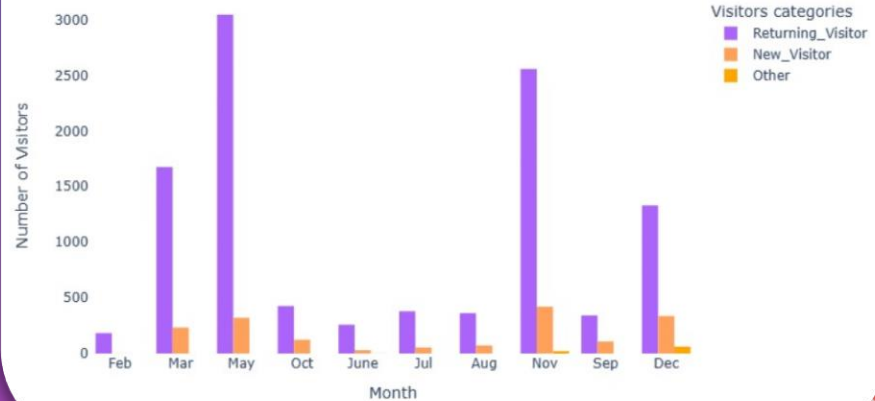


❑ 3 categories : Returning, New and Other visitors

❑ Only 10 months of the year represented

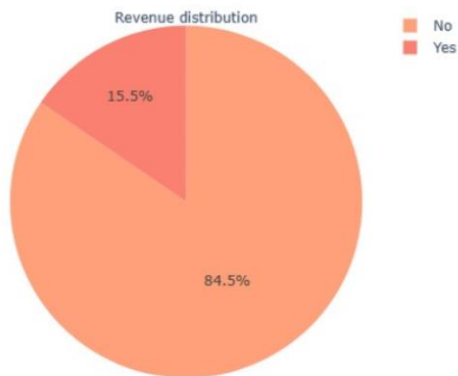
❑ 4 months on the top

Visitor Type Distribution by Month

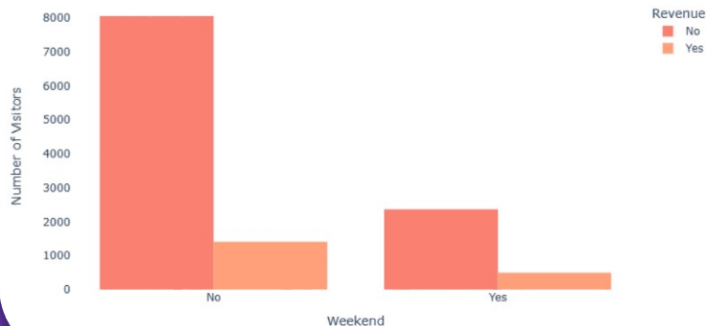




# Transactions study



Proportion of visits that led to a transaction during weekend



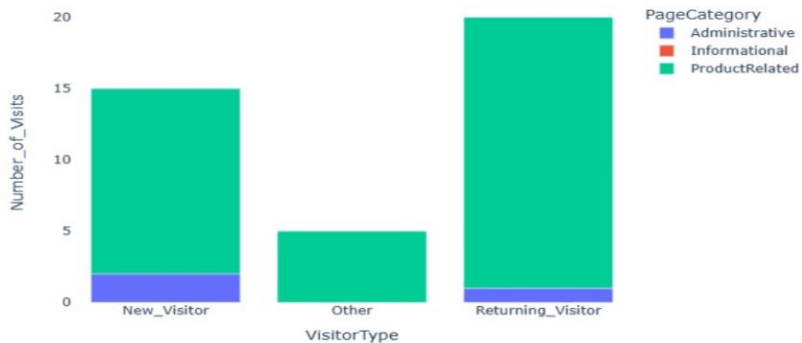
❑ Only 15,5% of visits led to a transaction.

❑ 21% of weekend visits led to transaction against 17% for weekday.



# Is there any difference on pages visits?

Median of Visits for Different Page Categories by visitors type



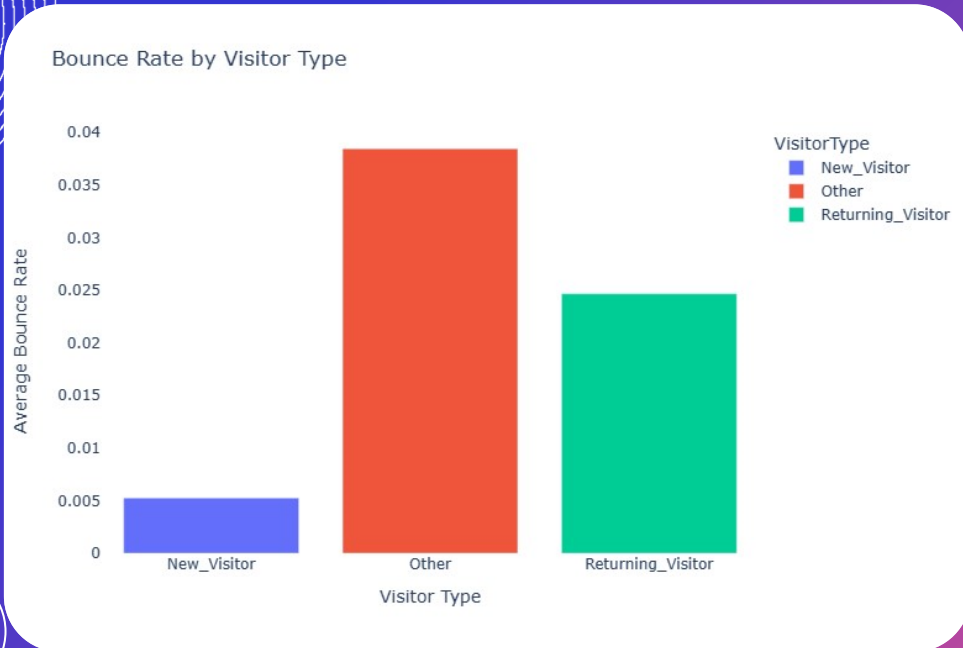
☐ New visitors tend to visit slightly more administrative pages.

☐ Returning visitors tend to visit more product-related pages.

Distribution of Administrative Pages by Visitor Type



# What for them to exit ?



❑ 4% for other

❑ 2.4% For returning visitors

❑ Only 0.5 for New ones



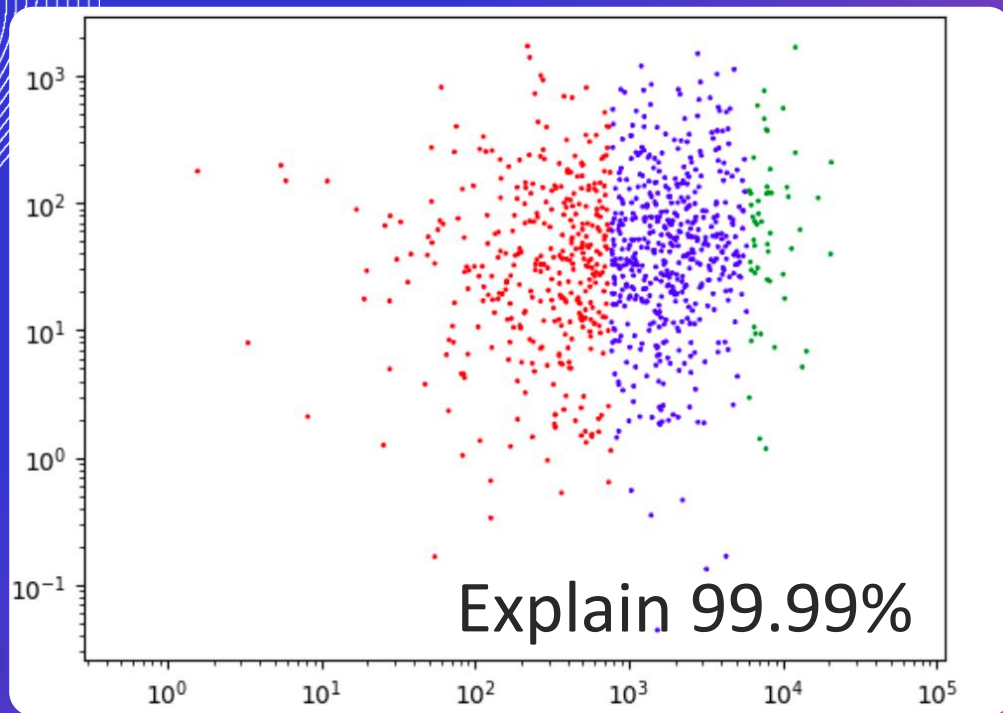
03

# Our process

Data pre-processing and modeling



# PCA



Clusters are  
not simple to  
define

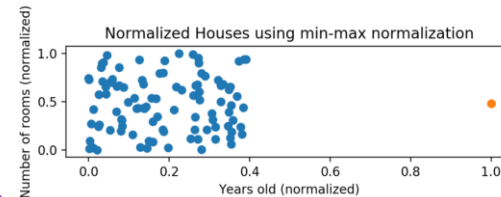
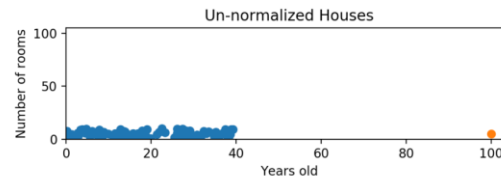
# Data Pre-processing used

## One-Hot Encoding

datagy.io

Island	Biscoe	Dream	Torgensen
Biscoe	1	0	0
Torgensen	0	0	1
Dream	0	1	0

Height	Height
Tall	0
Medium	1
Short	2



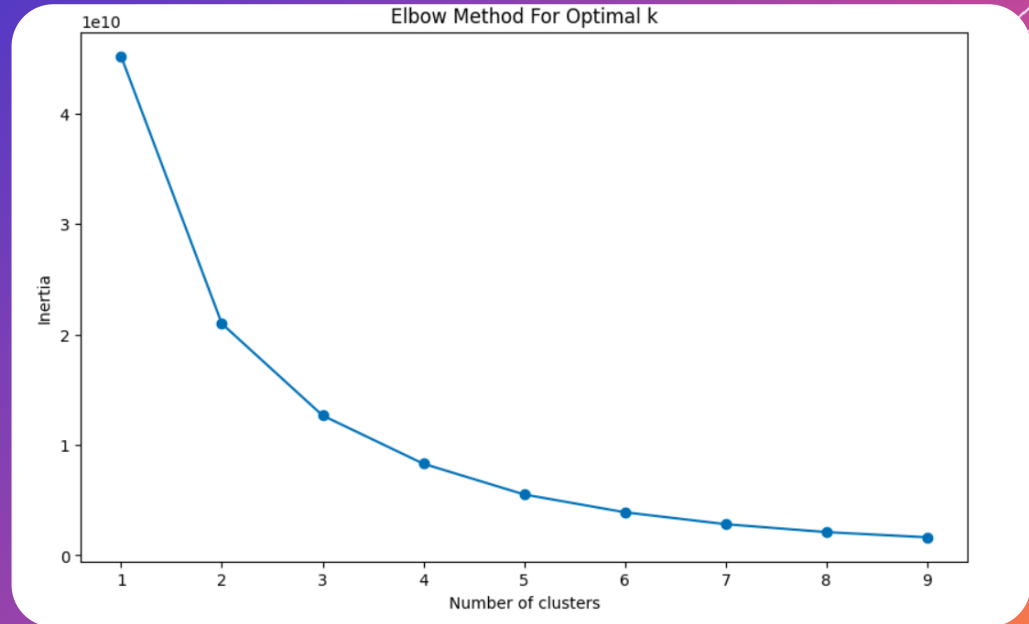


# Identify visitor who spend money

## K-Nearest Neighbors

Accuracy: 0.8503649635036497  
Precision: 0.7404564795116085  
Recall: 0.6231143552311436  
F1 Score: 0.6504117857907106  
Confusion Matrix:  
[[1981 74]  
[ 295 116]]

## KMeans



# PageValues prediction

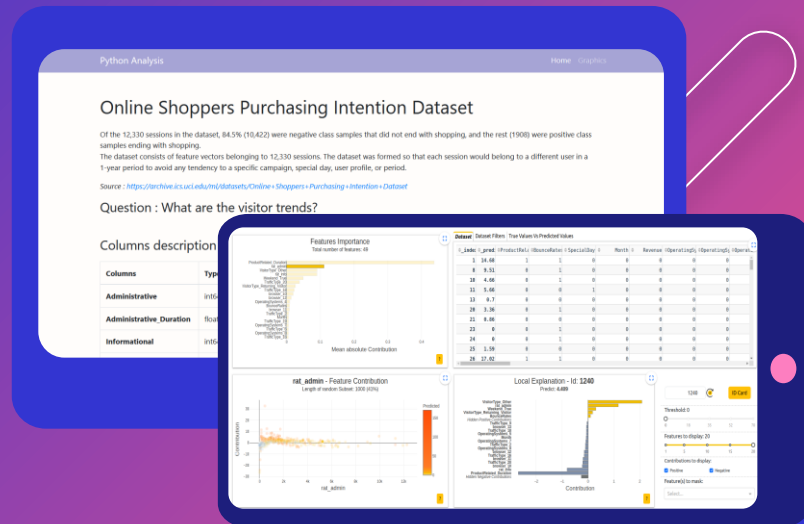
Grid Search Mean Error: 19.2

```
param_grid = {  
    'regression__C': [0.1,10],  
    'regression__gamma': [0.001,10],  
    'regression__kernel': ['rbf'],  
}
```

# Demonstration



Let's make a demo of the website and the Shapash report



# Conclusion

## Dataset



Data not very relevant  
Lack of information on  
some data.

A clean data, already  
encoded.



## Results

Different behaviors  
shown by the data viz  
regarding the category of  
visitors.

There are no significant  
clusters because of the  
uniformity of the dataset.





# Thanks!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for  
attribution