Raghad saleh alkhathran
Project proposal
3 October

Online Shoppers Intention

Now days data science become important aspect in different fields, as in Business field to understand the costumers. With so many potential sources of customer data, a foundational understanding of data science could help make sense of it.

Problem:

In this project I will analyze 'Online Shoppers Intention 'data to try improve the customers experience with the website, to gain more revenues

And answer these few questions:

- Is the weekend a factor to raise the revenue?
- the closeness of the site visiting time to a specific special day (e.g., Mother's Day, Valentine's Day) effect the revenue?
- What features has the strong correlation with revenue?

Dataset overview:

The dataset is an open source from Kaggle represent user who visited a website. The dataset consists of 10 numerical and 8 categorical attributes and 12,330 rows. some features represent the page that user had been visit and the duration the suer spend on it. And other represent the metrics measured by "Google Analytics". Also, the 'Special Day' feature one of the most important features when the user happen to visit the website in such a specific time like pre Mother's Day.

Feature:

Feature Name	Descriptions
Administrative	This is the number of pages of this type (administrative) that the user visited.
Administrative_Duration	This is the amount of time spent in this category of pages.
Informational	This is the number of pages of this type (informational) that the user visited.
Informational_Duration	This is the amount of time spent in this category of pages.
ProductRelated	This is the number of pages of this type (product related) that the user visited

Raghad saleh alkhathran
Project proposal
3 October

ProductRelated_Duration	This is the amount of time spent in this category of pages.
BounceRates	The percentage of visitors who enter the website through that page and exit without triggering any additional tasks.
ExitRates	The percentage of pageviews on the website that end at that specific page.
PageValues	The average value of the page averaged over the value of the target page and/or the completion of an eCommerce
SpecialDay	This value represents the closeness of the browsing date to special days or holidays (eg Mother's Day or Valentine's day) in
Month	In which month did he visit the site?
OperatingSystems	What operating system the visitor use
Browser	What browser the visitor use
Region	The visitor from any region ?
TrafficType	traffic type for the visitor
VisitorType	visitor type as returning or new visitor
Weekend	Boolean value indicating whether the date of the visit is weekend
Revenue	Boolean value indicating if the visitor paid after visiting the website, False if not

Data Link : https://www.kaggle.com/imakash3011/online-shoppers-purchasing-intention-dataset

Tools:

I will work to clean the dataset to find any null value or inconsistent in the data and other cleaning process, then I can visualize the data and learn more about it. The Plan is to predict 'Revenue' attribute and classified as true when the user paid after visiting the website, False if not. Also, to answer the previous questions, I need to study how the user behaving by using the dataset. additionally, split the data into train set and test set to training and test the classification model to predict the ‘Revenue’. By using python and it’s libraries (e.g. pandas and matplotlib)

Raghad saleh alkhathran
Project proposal
3 October

Jupyter Notebook: Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations.

Python libraries might be used:

- NumPy: the fundamental library for scientific computing in Python on which Pandas was built.
- Pandas: to do data manipulation and analysis.
- Matplotlib: for data visualization
- Seaborn: data visualization library for statistical graphics plotting in Python.

What I expected after study is the revenue raise in the weekend and close to special days, additionally the result it could help to identifying best time for offers. The next step after the proposal is the MVP, which will contain the data after cleaning and EDA (Exploratory Data Analysis) to discover and learn more about data.

Raghad saleh alkhathran
Project proposal
3 October

References:

- https://online.hbs.edu/blog/post/what-is-data-science
- https://jupyter.org
- https://www.datacamp.com/community/blog/python-pandas-cheat-sheet?utm_source=adwords_ppc&utm_campaignid=12492439802&utm_adgroupid=122563403481&utm_device=c&utm_keyword=panda%20package%20python&utm_matchtype=b&utm_network=g&utm_adposition=&utm_creative=504158804617&utm_targetid=aud-1282405912307:kwd-614516587896&utm_loc_interest_ms=&utm_loc_physical_ms=9077037&gclid=CjwKCAjwqeWKBhBFEiwABo_XBpk2ARUn0ShWHZP8Ginyu4OJPPd5SD2O-ajSgdUwx_v5WUCPy6CTQRoCQrYQAvD_BwE
- https://www.mygreatlearning.com/blog/seaborn-tutorial/