Data Description for E-commerce python EDA & Hypothesis testing

Shopping dataset

Column Descriptions:

- **Administrative:** This is the number of pages of this type (administrative) that the user visited. Eg: account, addresses, cart, orders etc
- **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.
- **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 transaction.
  More information about how this is calculated
- **SpecialDay:** This value represents the closeness of the browsing date to special days or holidays (eg Mother's Day or Valentine's day) in which the transaction is more likely to be finalized. More information about how this value is calculated below.
- **Month:** Contains the month the pageview occurred, in string form.
- **OperatingSystems:** An integer value representing the operating system that the user was on when viewing the page.

- **Browser:** An integer value representing the browser that the user was using to view the page.

- **Region:** An integer value representing which region the user is located in.

- **TrafficType:** An integer value representing what type of traffic the user is categorized into.

- **VisitorType:** A string representing whether a visitor is New Visitor, Returning Visitor, or Other.

- **Weekend:** A boolean representing whether the session is on a weekend.

- **Revenue:** A boolean representing whether or not the user completed the purchase.

The 'Revenue' attribute can be used as the class label. Of the 12,330 sessions in the dataset, 84.5% (10,422) were negative class samples that did not end with shopping, and the rest (1908) were positive class samples ending with shopping.

"Administrative", "Administrative Duration", "Informational", "Informational Duration", "Product Related" and "Product Related Duration" represent the number of different types of pages visited by the visitor in that session and total time spent in each of these page categories.

The values of these features are derived from the URL information of the pages visited by the user and updated in real time when a user takes an action, e.g. moving from one page to another.

The "Bounce Rate", "Exit Rate" and "Page Value" features represent the metrics measured by "Google Analytics" for each page in the e-commerce site. The value of "Bounce Rate" feature for a web page refers to the percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session.

The value of "Exit Rate" feature for a specific web page is calculated as for all pageviews to the page, the percentage that were the last in the session. The "Page Value" feature represents the average value for a web page that a user visited before completing an e-commerce transaction.

The "Special Day" feature indicates the closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with transaction.

The value of this attribute is determined by considering the dynamics of e-commerce such as the duration between the order date and delivery date. For example, for Valentina's day, this value takes a nonzero value between February 2 and February 12, zero before and after this date unless it is close to another special day, and its maximum value of 1 on February 8.

The dataset also includes operating system, browser, region, traffic type, visitor type as returning or new visitor, a Boolean value indicating whether the date of the visit is weekend, and month of the year.

Campaign dataset

Column Details:

- ID: Customer's Unique Identifier

- Year_Birth: Customer's Birth Year

- Education: Customer's education level (Graduation, Master, PhD, 2n Cycle(Diploma), Basic)

- Marital_Status: Customer's marital status

- Income: Customer's yearly household income

- Kidhome: Number of children in customer's household

- Teenhome: Number of teenagers in customer's household

- Dt_Customer: Date of customer's enrollment with the company

- Recency: Number of days since customer's last purchase

- MntWines: Amount spent on wine in the last 2 years

- MntFruits: Amount spent on fruits in the last 2 years

- MntMeatProducts: Amount spent on meat in the last 2 years

- MntFishProducts: Amount spent on fish in the last 2 years

- MntSweetProducts: Amount spent on sweets in the last 2 years

- MntGoldProds: Amount spent on gold in the last 2 years

- NumDealsPurchases: Number of purchases made with a discount

- NumWebPurchases: Number of purchases made through the company's web site

- NumCatalogPurchases: Number of purchases made using a catalogue

- NumStorePurchases: Number of purchases made directly in stores

- NumWebVisitsMonth: Number of visits to company's web site in the last month

- AcceptedCmp1: 1 if customer accepted the offer in the 1st campaign, 0 otherwise (Target variable)

- AcceptedCmp2: 1 if customer accepted the offer in the 2nd campaign, 0 otherwise (Target variable)
- AcceptedCmp3: 1 if customer accepted the offer in the 3rd campaign, 0 otherwise (Target variable)
- AcceptedCmp4: 1 if customer accepted the offer in the 4th campaign, 0 otherwise (Target variable)
- AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise (Target variable)
- Complain: 1 if customer complained in the last 2 years, 0 otherwise
- Country: Customer's location

This is a CSV file of observations (customers) with variables related to marketing data. More specifically, the variables provide insights about:

- Customer profiles
- Products purchased
- Campaign success (or failure)
- Channel performance