

United States E-Commerce records 2020

DATA

Source of Data: The dataset is available on an open-source platform **Kaggle** and can be download from the platform

Dataset - Click [Link](#)

Collection Methodology - Web Scrapping

Data Content – The content contains information about United States E-Commerce records 2020 and contains state,

- Order date
- Row id
- Order ID
- Ship mode
- Customer id
- Segment
- Country
- State
- Postal code
- City
- Region
- Product ID
- Category
- Sub-Category
- Sales
- Quantity
- Discount
- Profit

The dataset contains 19 columns and 3312 rows, Quality check, integrity and wrangling was done as followed

Check for null – no null value

Check for mix data type – no mix data type

Check for duplicate – no duplicate found

When the statistics was done using describe () the profit min value contains -ve value and the further investigation on the profit column shows 620 rows contains -ve value and this is an error during computation.

Futhermore, the absolute function (abs) was used to fix the -ve rows

Column details

Column	Quatitative/Qualitative	Decrete/ Continuous	Norminal/ordinal/Binary
Order Date	Qualitative		Norminal
Row ID	Qualitative		Norminal
Order ID	Qualitative		Norminal
Ship Mode	Qualitative		Ordinal
Customer ID	Qualitative		Norminal
Segment	Qualitative		Norminal
Country	Qualitative		Norminal
City	Qualitative		Norminal
State	Qualitative		Norminal
Postal Code	Qualitative		Norminal
Region	Qualitative		Norminal
Product ID	Qualitative		Norminal
Category	Qualitative		Norminal
Sub-Category	Qualitative		Norminal
Product Name	Qualitative		Norminal
Sales	Quatitative	Continuous	
Quantity	Quatitative	Decrete	
Discount	Quatitative	Continuous	
Profit	Quatitative	Continuous	

Limitation and ethics

1. Limitation: The Profit column contains -ve value, this might be due to manual computation
2. Ethical issue: The data contains no personal Information or indiscriminates variable,so no PLA Security is required.

List of Questions to answers

Analysis

- Which of the category or sub-category generate more revenue
- How do the sales vary over different months
- What category or sub category are the best-sellers in terms of quantity sold
- What is the type of category or sub-category are mostly order by each region
- Which of ship mode receive more discount
- Is there any correlation between shipping mode and discount
- Is there any correlation between quantity order and discount
- Do discount led to more sales
- Which of the state order more of the product

Hypothesis

If a state in a region orders more quantity, then so the revenue generated from the quantity is higher.

If region has more order in quantity, then the revenue generated from sales in the region will be higher.