eCommerce Business

eCommerce Behavior Data from Multi Category Store

Outline

Understanding eCommerce Business

Data Preprocessing

Exploratory Data Analysis


Understanding
the eCommerce
Business

eCommerce Business

eCommerce Business is important to make a highly profit. In 2020, over two billion people from seven billion population in the world purchased online products and services.

Analysing customer behaviour pattern, we can increase customer value and therefore spend by content personalisation and optimisation. Based on this exploratory data analysis, a profile of the customer is created and grouped into segments by their digital footprint and interests.

Data Behavioral

See how customers interact with the platform. Like clicks per page. It can also be a flow from the customer to the purchase of products.

Observation of a variable in relation to other variables from one point to another point.

How does the platform engage customers.

What information are we looking for?

1. How is the trend of daily visiting during October?
2. How is the trend of daily visitors during October?
3. How is the trend of daily conversation rates during October?
4. How is the analysis for the category, customer and event typer?


Data
Preprocessing

What is data preprocessing?

Data preprocessing is the process (**transformation**) of raw data into an understandable format. Raw data is usually incomplete (**missing value**), inconsistent, noisy (**many errors**) and very extreme value (**outlier**)

#Data Cleansing

Remove Duplicates

By default, it removes duplicate rows based on all columns.

We can use this command to remove the duplicate.

```
#Remove the duplicates Data
df.drop_duplicates(inplace=True)
```

Drop Column

We can drop the category_id column, because we don't use it in the EDA.

We can use this command to drop the column.

```
df.drop(columns=['category_id'],inplace=True)
```

#Data Cleansing

Handling Missing Value

```
event_time          0.00
event_type          0.00
product_id          0.00
category_code      31.85
brand              14.42
price               0.00
user_id             0.00
user_session        0.00
dtype: float64
```

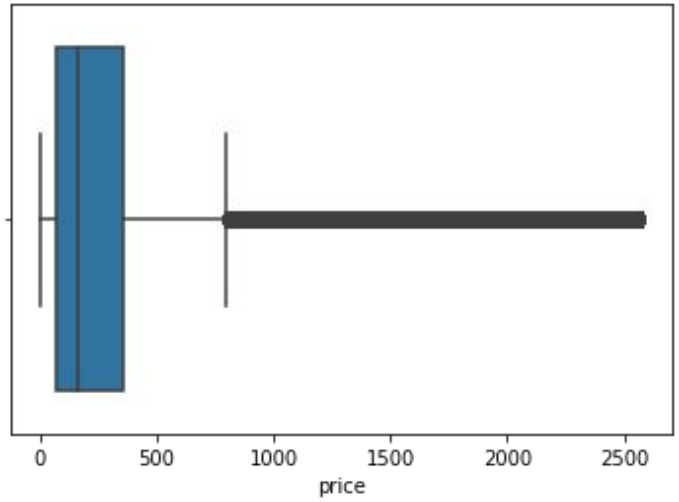
There are 31.85% of missing value in category_code and 14.42% of missing value in brand.

Here we assume that there are product that do not have a brand, so they are replaced with ‘nobrand’.

We don't eliminate products that don't have categories because we want to see the overall data visualization, so they are replaced with ‘nocategory’.

#Data Cleansing

Check the Outliers



There are **3,668,541** outlier price data. We don’t need to remove the outlier because that is a true outlier.

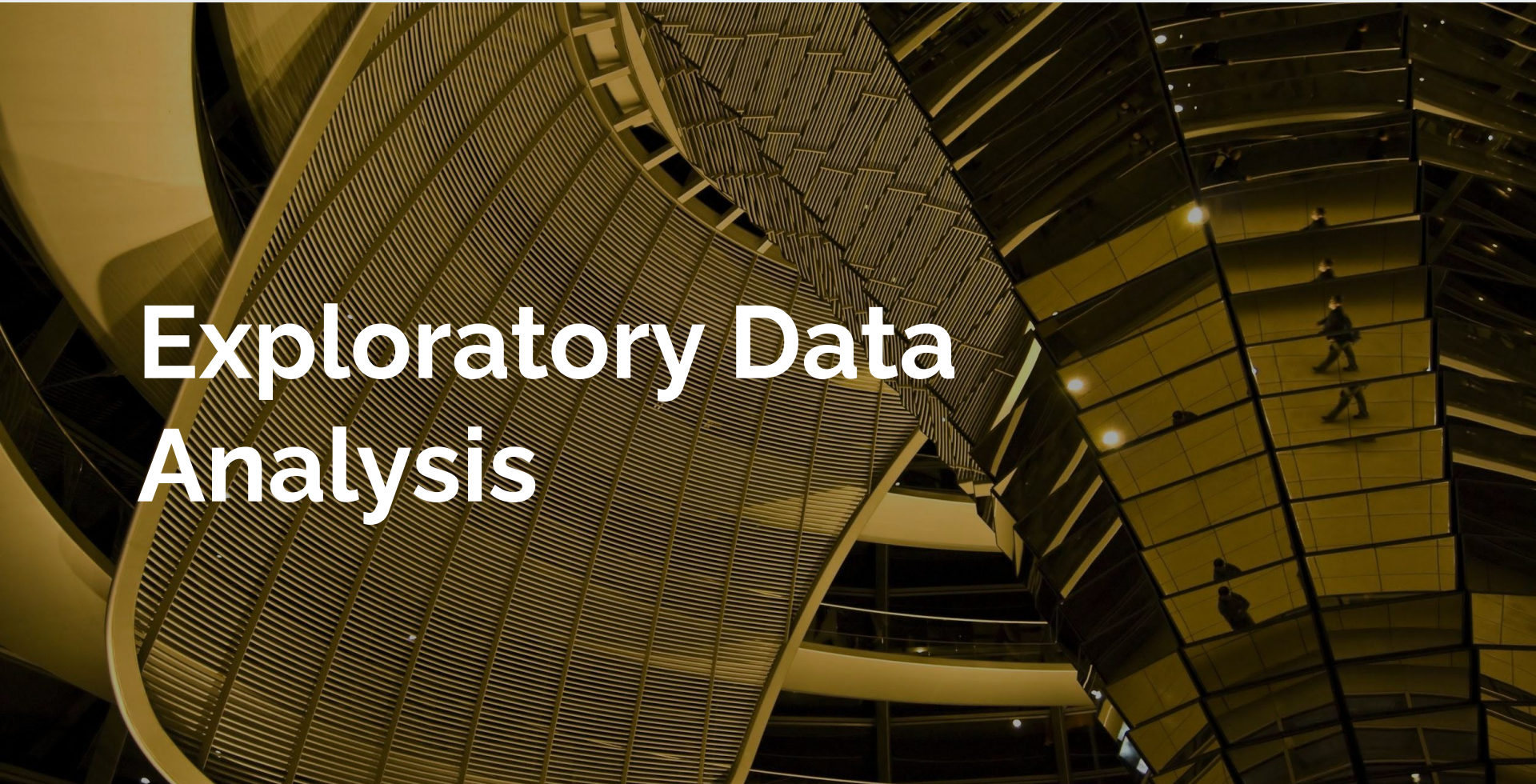
#Data Cleansing

Handling Invalid Data Types

```
event_time        object
event_type        object
product_id         int64
category_code     object
brand             object
price            float64
user_id            int64
user_session      object
dtype: object
```

Event_time must be a datetime data type, so we need to change it.

Also we can change data type of event_type and category_code to category data type, this is for easier visualize data in EDA.


Exploratory Data
Analysis

Daily Visiting Trend

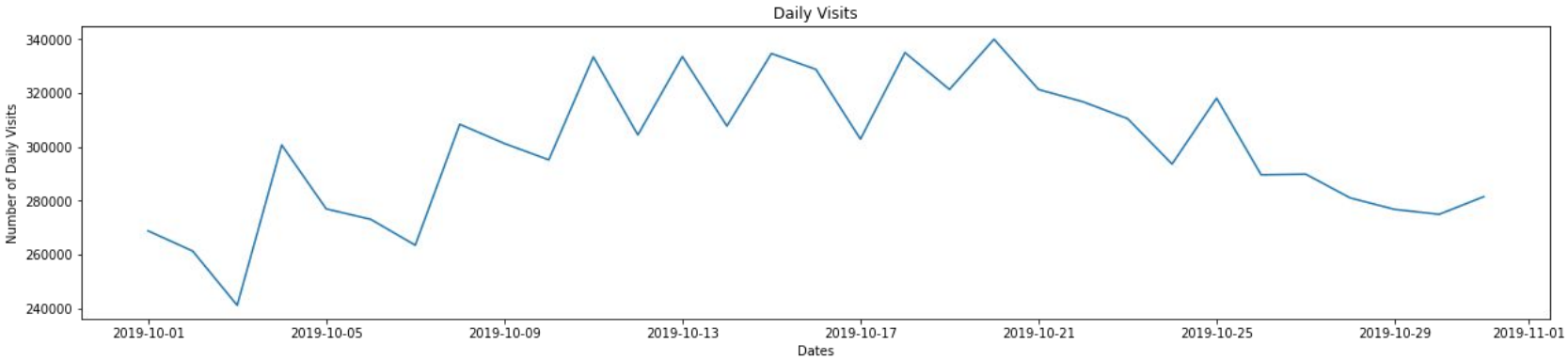
01

Average daily visits is **299,510.55.** We can see the visit statistics by date:

```
-------------------------------------------------
Visit Statistics by Dates
-------------------------------------------------
            count       mean            std  ...       50%        75%       max
event_time                                   ...
Friday        4.0  321772.50  16017.261304  ...  325731.0  333799.50  334977.0
Monday        4.0  293353.00  26033.151442  ...  294346.5  311089.50  321282.0
Saturday      4.0  298041.75  19140.215435  ...  296976.5  308605.50  321300.0
Sunday        4.0  309080.00  32756.353552  ...  311671.5  335116.75  339943.0
Thursday      5.0  282820.60  24557.356776  ...  293587.0  295142.00  302842.0
Tuesday       5.0  301032.40  27672.766817  ...  308348.0  316706.00  334648.0
Wednesday     5.0  295314.60  27194.956762  ...  301219.0  310453.00  328743.0
```

Daily Visiting Trend

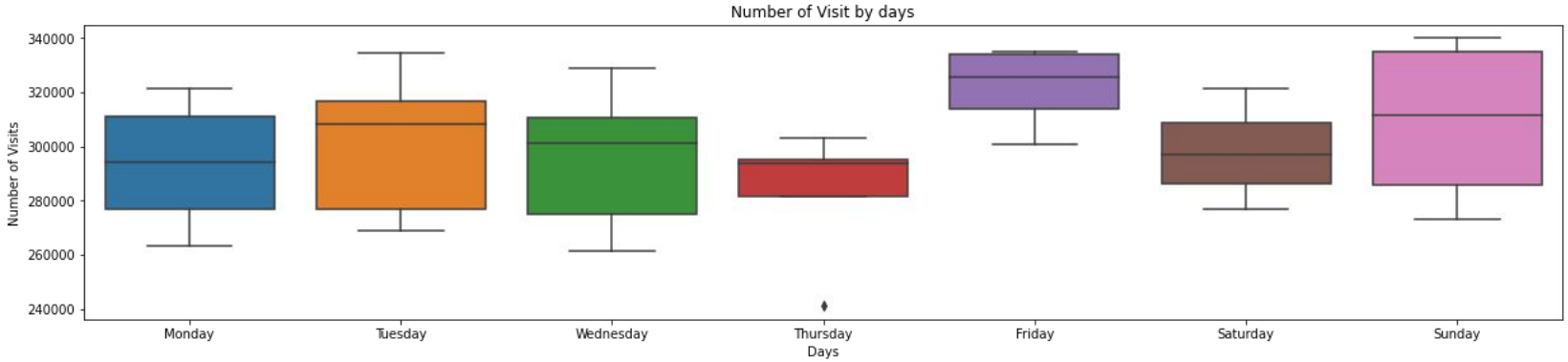
01



There is a fluctuating trend of daily visits. The highest visit occurred in the mid-month period.

Daily Visiting Trend

01



The highest visits during October occur on Sundays and the lowest on Thursdays,

Daily Visitors Trend

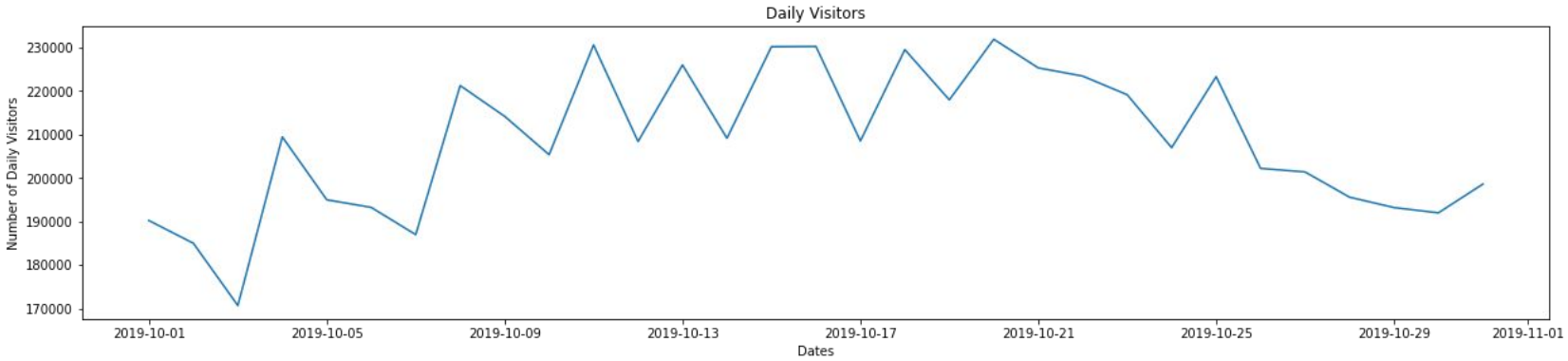
02

Average daily visits is **208,829.77**. We can see the visitor statistics by date:

```
--------------------------------------------------
Visitor Statistics by Dates
--------------------------------------------------
             count       mean           std  ...       50%        75%       max
event_time                                   ...
Friday         4.0  223166.00   9715.295466  ...  226360.5  229730.50  230533.0
Monday         4.0  204218.00  16732.498877  ...  202326.0  213134.00  225269.0
Saturday       4.0  205851.25   9731.790975  ...  205263.5  210746.25  217920.0
Sunday         4.0  213090.00  18707.438325  ...  213650.5  227418.75  231849.0
Thursday       5.0  197994.80  15736.888740  ...  205321.0  206937.00  208477.0
Tuesday        5.0  211615.80  18526.041245  ...  221204.0  223384.00  230135.0
Wednesday      5.0  208073.80  18982.955637  ...  214140.0  219100.00  230199.0
```

Daily Visitors Trend

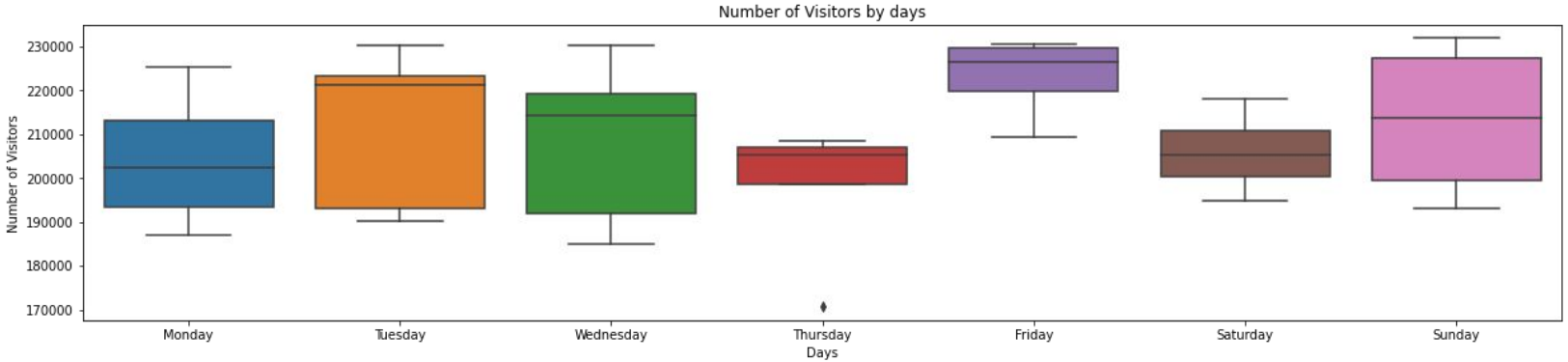
02



There is a fluctuating trend of daily visitors. The highest visitors occurred in the mid-month period.

Daily Visitors Trend

02



The highest visitors during October occur on Tuesday and the lowest on Thursdays.

Conversion Rates Trend

03

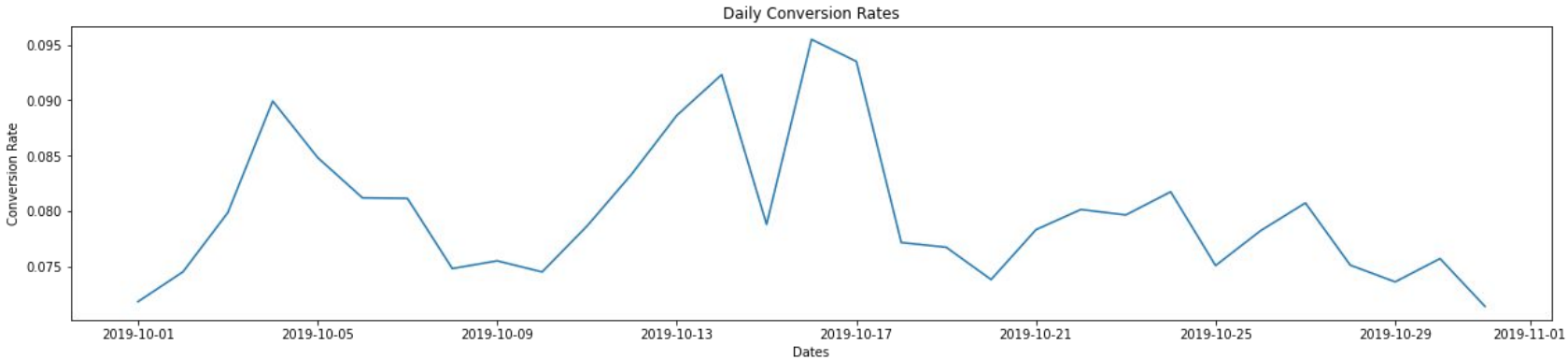
Average daily visits is **0.079887**. We can see the conversion rate statistics by date:

```
-------------------------------------------------
Conversion Rates Statistics by Dates
-------------------------------------------------
```

	count	mean	std	...	50%	75%	max
event_time				...			
Friday	4.0	0.080212	0.006645	...	0.077911	0.081475	0.089935
Monday	4.0	0.081730	0.007474	...	0.079739	0.083942	0.092316
Saturday	4.0	0.080791	0.003913	...	0.080794	0.083730	0.084835
Sunday	4.0	0.081099	0.006051	...	0.080967	0.083055	0.088637
Thursday	5.0	0.080207	0.008502	...	0.079868	0.081744	0.093504
Tuesday	5.0	0.075845	0.003512	...	0.074821	0.078802	0.080147
Wednesday	5.0	0.080181	0.008782	...	0.075717	0.079658	0.095494

Conversion Rates Trend

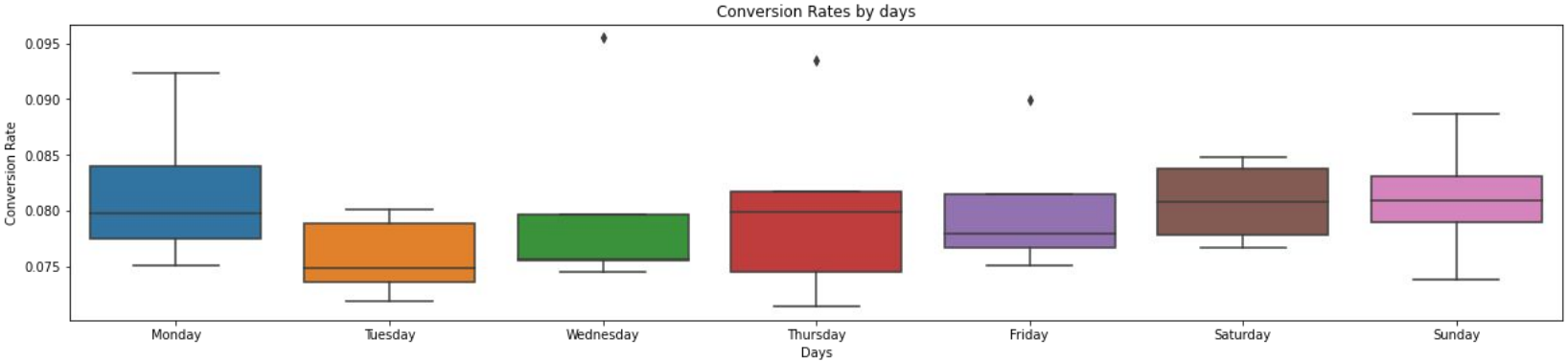
03



There is a fluctuating trend of daily conversion rates. The highest conversion rates occurred in the mid-month period and tends to decrease at the end of the month.

Conversion Rates Trend

03



Conversion rate on each day during October tends to be the same, which is the highest one is on Thursday.

Customer Analysis

04

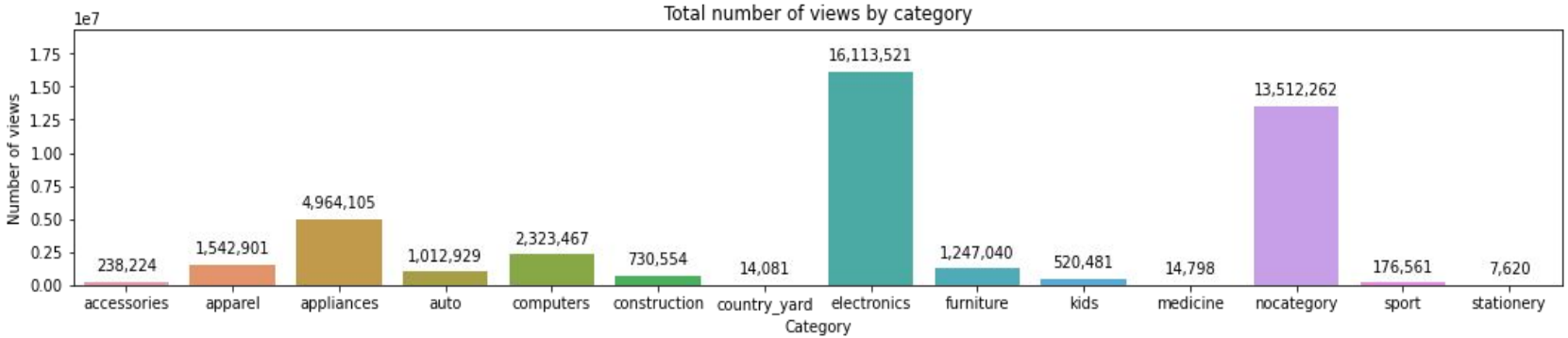
There are **347,118** customers who purchased in October and there are **131,408** repeat customers who bought more than once. This indicates that **37.86%** visitors have successfully directed repeat purchases. We can see the summary statistics of top 10% customers and regular customers.

```
Top 10% customers Purchase Amount-Descriptive Statistics
--------------------------------------------------
count     34781.000000
mean       3730.465088
std        5620.445256
min        1418.050000
25%        1741.340000
50%        2335.420000
75%        3766.790000
max      265569.520000
Name: total_sales, dtype: float64
--------------------------------------------------
Total sales: 229,933,213
Total sales of top 10% customers: 129,749,306
```

```
Regular customers Purchase Amount-Descriptive Statistics
--------------------------------------------------
count    312337.000000
mean        320.755807
std         314.502807
min           0.880000
25%          94.210000
50%         204.870000
75%         442.180000
max        1418.040000
Name: total_sales, dtype: float64
--------------------------------------------------
Total sales of regular customers: 100,183,906
```

Category Analysis

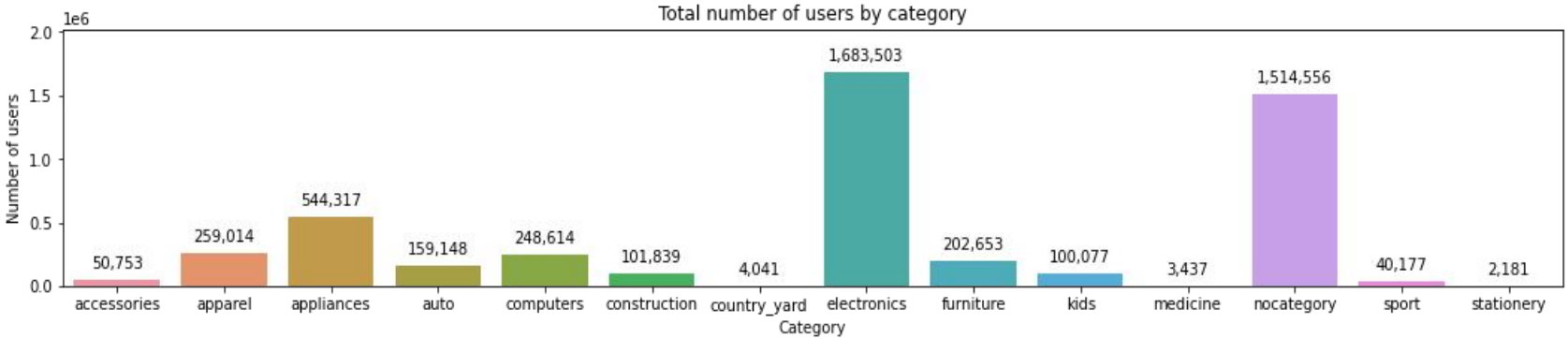
05



Electronics is at the top of the most viewed category with **16,113,521** views.

Category Analysis

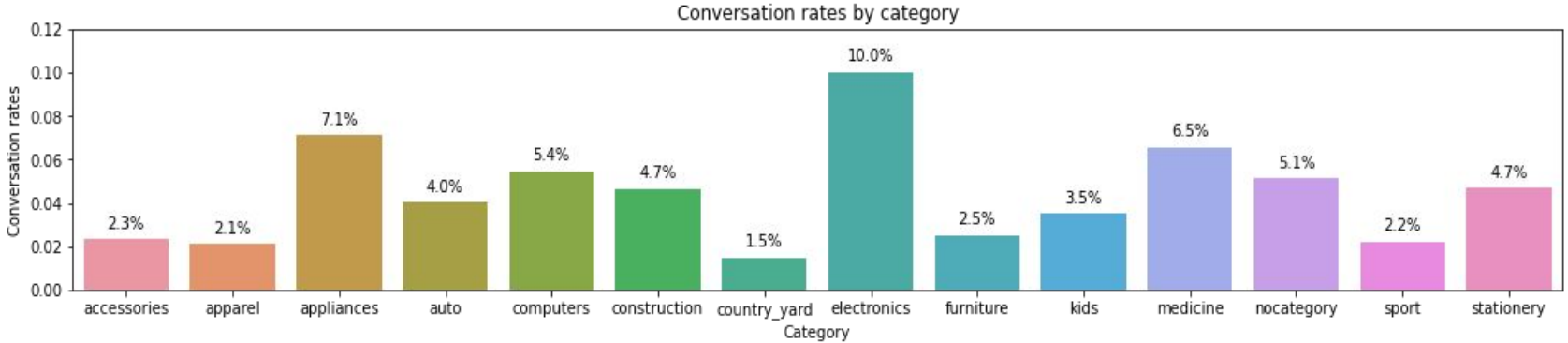
05



Electronics is at the top of the most users category with **1,683,503** users.

Category Analysis

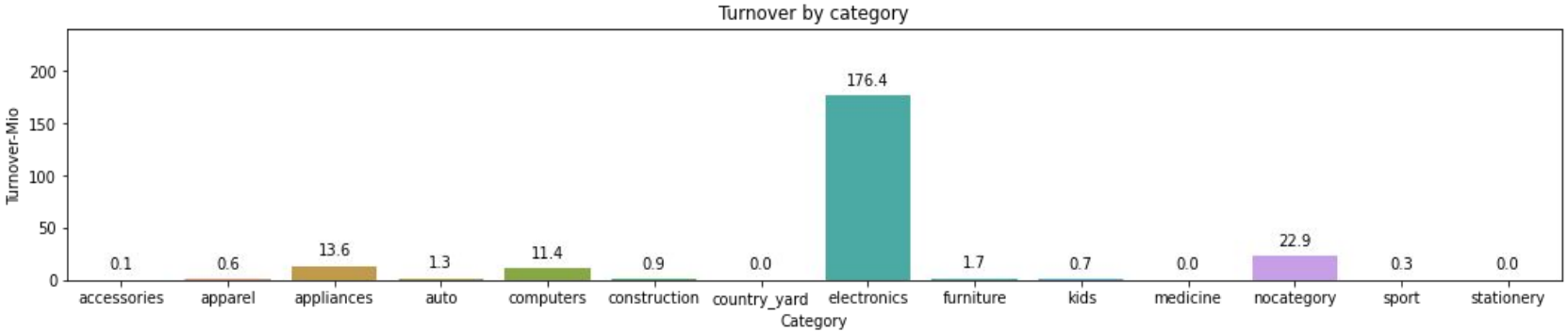
05



Electronics is at the top of the most conversion rates category with **10%** conversion rates.

Category Analysis

05

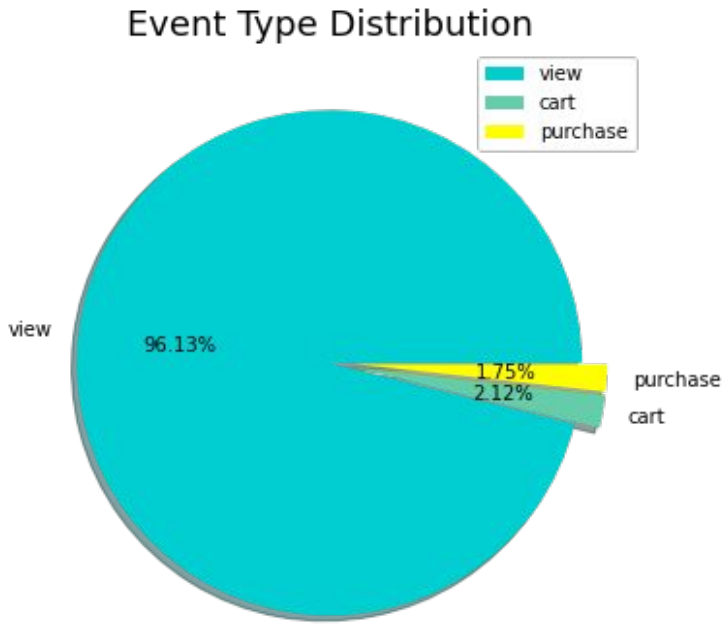


Electronics is at the top of the most turnover category with **176,4** turnover.

Event Type Analysis

06

There are three categories of event types. The most event type is view with **96.13%.**



Thank you.

There are three categories of event types. The most event type is view with **96.13%**.