```
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go
```

```
df=pd.read_csv("ecommerce_customer_data.csv")
df.head() #view dataset
```

	User_ID	Gender	Age	Location	Device_Type	Product_Browsing_Time	Total_Pages_Viewed	Items_Added_to_Cart	Total_Purchases
0	1	Female	23	Ahmedabad	Mobile	60	30	1	0
1	2	Male	25	Kolkata	Tablet	30	38	9	4
2	3	Male	32	Bangalore	Desktop	37	13	5	0
3	4	Male	35	Delhi	Mobile	7	20	10	3
4	5	Male	27	Bangalore	Tablet	35	20	8	2

```
df.describe() #numeric summary
```

	User_ID	Age	Product_Browsing_Time	Total_Pages_Viewed	Items_Added_to_Cart	Total_Purchases
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	250.500000	26.276000	30.740000	27.182000	5.150000	2.464000
std	144.481833	5.114699	15.934246	13.071596	3.203127	1.740909
min	1.000000	18.000000	5.000000	5.000000	0.000000	0.000000
25%	125.750000	22.000000	16.000000	16.000000	2.000000	1.000000
50%	250.500000	26.000000	31.000000	27.000000	5.000000	2.000000
75%	375.250000	31.000000	44.000000	38.000000	8.000000	4.000000
max	500.000000	35.000000	60.000000	50.000000	10.000000	5.000000

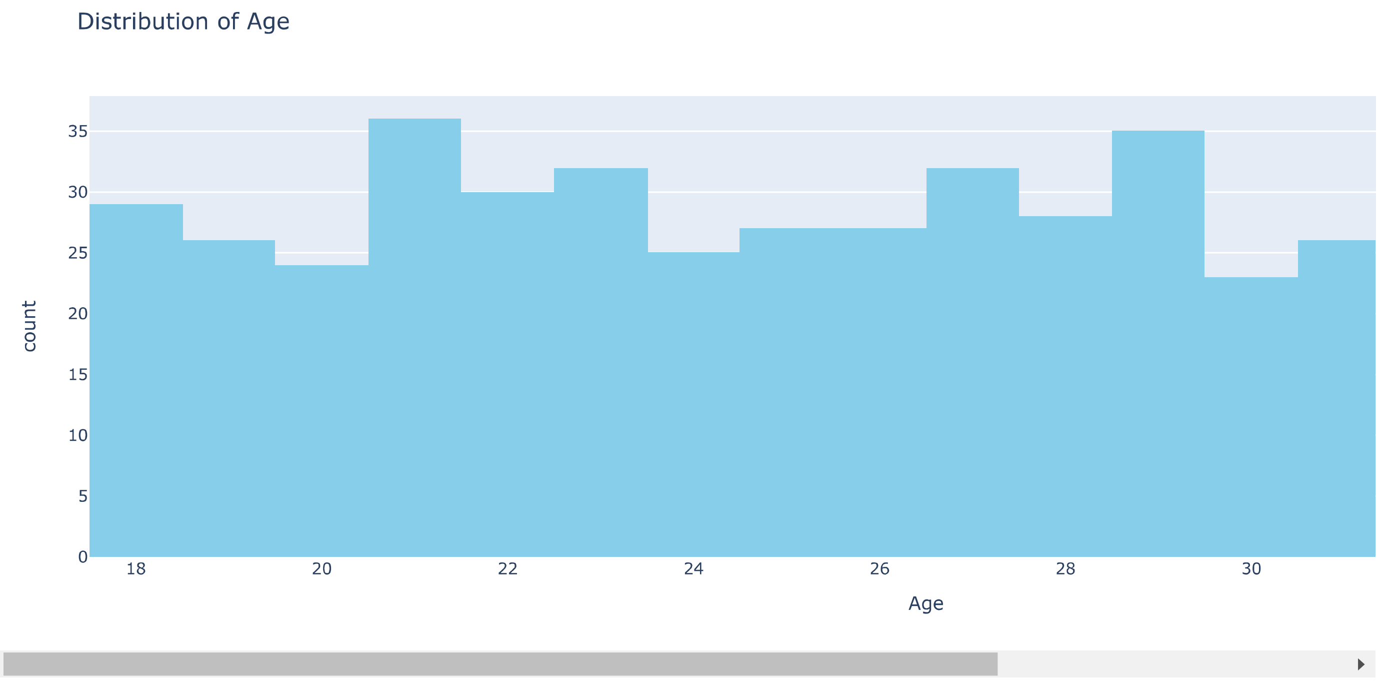
```
#categorical_summary
df.describe(include='object')
```

	Gender	Location	Device_Type
count	500	500	500
unique	2	8	3
top	Male	Kolkata	Mobile
freq	261	71	178

```
#checking null values
df.isnull().sum()
```

```
User_ID                  0
Gender                   0
Age                      0
Location                 0
Device_Type              0
Product_Browsing_Time    0
Total_Pages_Viewed       0
Items_Added_to_Cart      0
Total_Purchases          0
dtype: int64
```

```
#check the distribution of 'Age' in dataset
fig=px.histogram(df,x='Age',title='Distribution of Age',nbins=20,color_discrete_sequence=['skyblue'])
fig.show()
```



```
df['Gender'].value_counts() #gender counts
```

```
Male      261
Female    239
Name: Gender, dtype: int64
```

Lets Check cor-relation

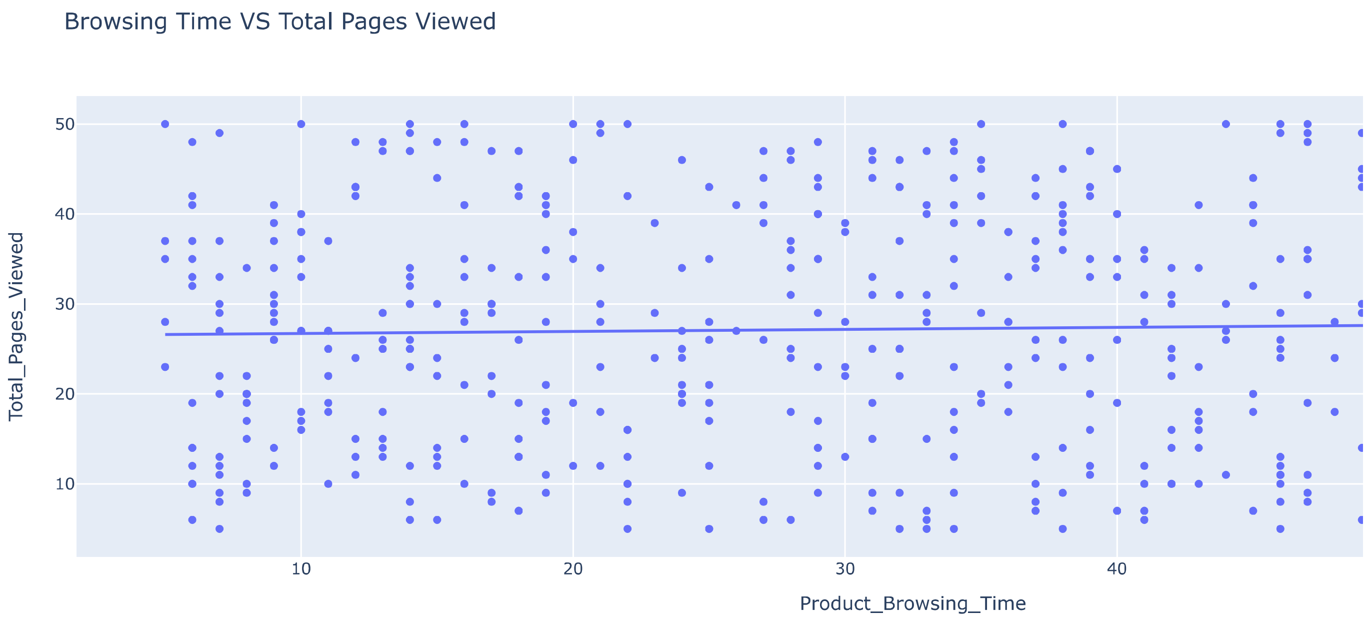
```
df.corr(numeric_only=True)
```

	User_ID	Age	Product_Browsing_Time	Total_Pages_Viewed	Items_Added_to_Cart	Total_Purchases
User_ID	1.000000	-0.054237	-0.058246	-0.076933	-0.024719	-0.019974
Age	-0.054237	1.000000	-0.055870	0.015823	-0.031033	-0.042094
Product_Browsing_Time	-0.058246	-0.055870	1.000000	0.027841	0.017531	-0.023239
Total_Pages_Viewed	-0.076933	0.015823	0.027841	1.000000	0.024570	0.013630
Items_Added_to_Cart	-0.024719	-0.031033	0.017531	0.024570	1.000000	0.007978
Total_Purchases	-0.019974	-0.042094	-0.023239	0.013630	0.007978	1.000000

Analyzing Customer Behaviour

> Now, let’s have a look at the relationship between the product browsing time and the total pages viewed:

```
fig=px.scatter(df,x='Product_Browsing_Time',y='Total_Pages_Viewed',title='Browsing Time VS Total Pages Viewed',trendline='ols')
fig.show()
```



The above scatter plot shows no consistent pattern or strong association between the time spent browsing products and the total number of pages viewed. It indicates that customers are not necessarily exploring more pages if they spend more time on the website, which might be due to various factors such as the website design, content relevance, or individual user preferences.

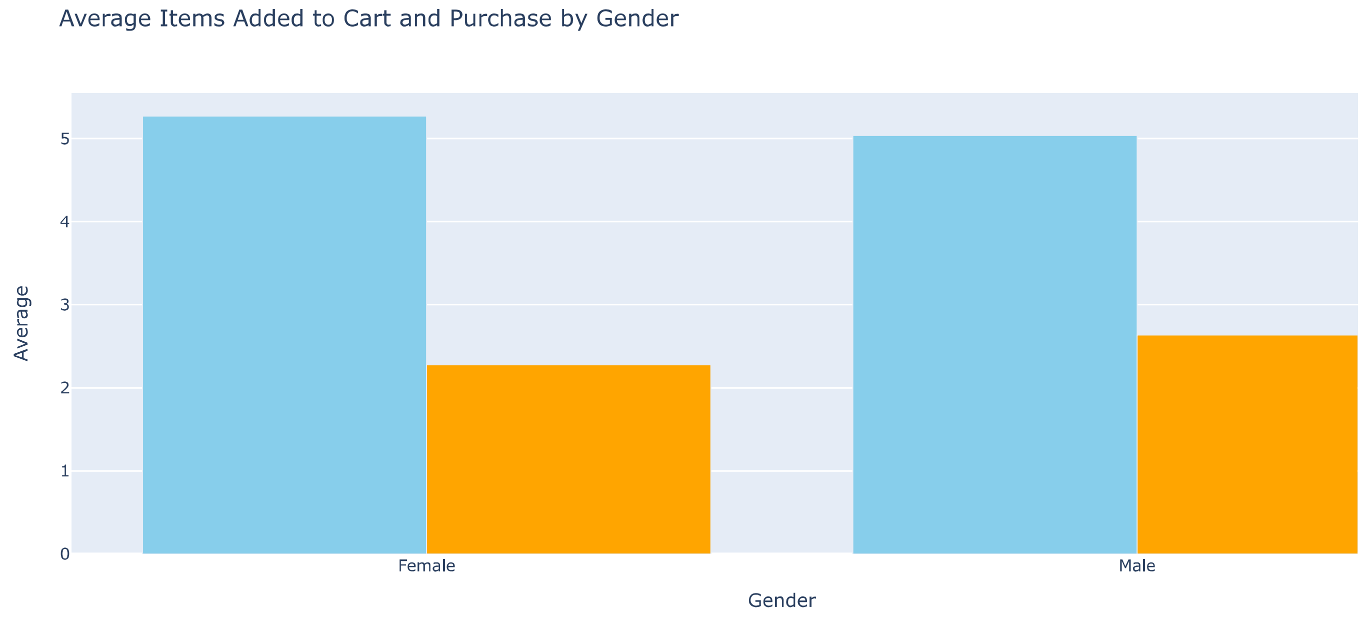
```
import plotly.express as px

# Assuming df is your DataFrame
gender_grouped = df.groupby('Gender').agg({'Items_Added_to_Cart': 'mean', 'Total_Purchases': 'mean'}).reset_index()
gender_grouped.columns = ['Gender', 'Average_Items_Added_to_Cart', 'Average_Purchase']

fig = px.bar(gender_grouped, x='Gender', y=['Average_Items_Added_to_Cart', 'Average_Purchase'],
             title='Average Items Added to Cart and Purchase by Gender',
             labels={'value': 'Average', 'variable': 'Metric', 'Gender': 'Gender'},
             color_discrete_sequence=['skyblue', 'orange'],
             barmode='group')

fig.update_layout(xaxis_title='Gender', yaxis_title='Average', legend_title='Metric')

fig.show()
print(gender_grouped)
```

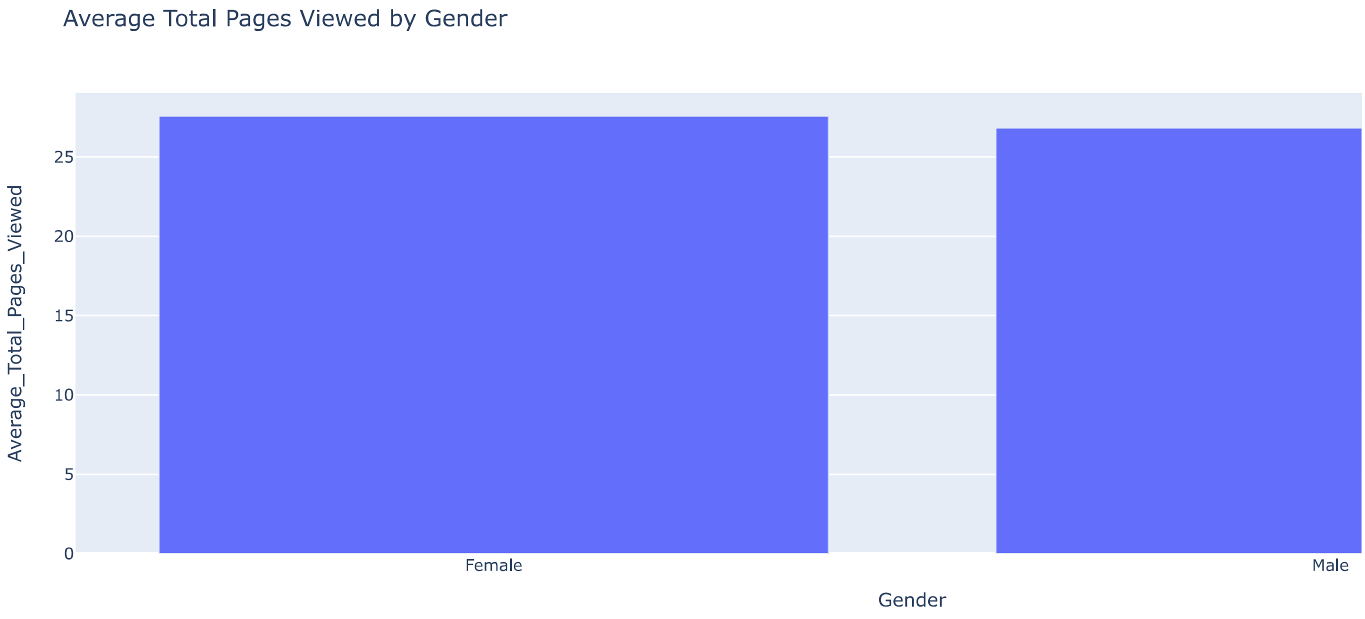


```
   Gender  Average_Items_Added_to_Cart  Average_Purchase
0  Female                     5.271967          2.276151
1    Male                     5.038314          2.636015
```

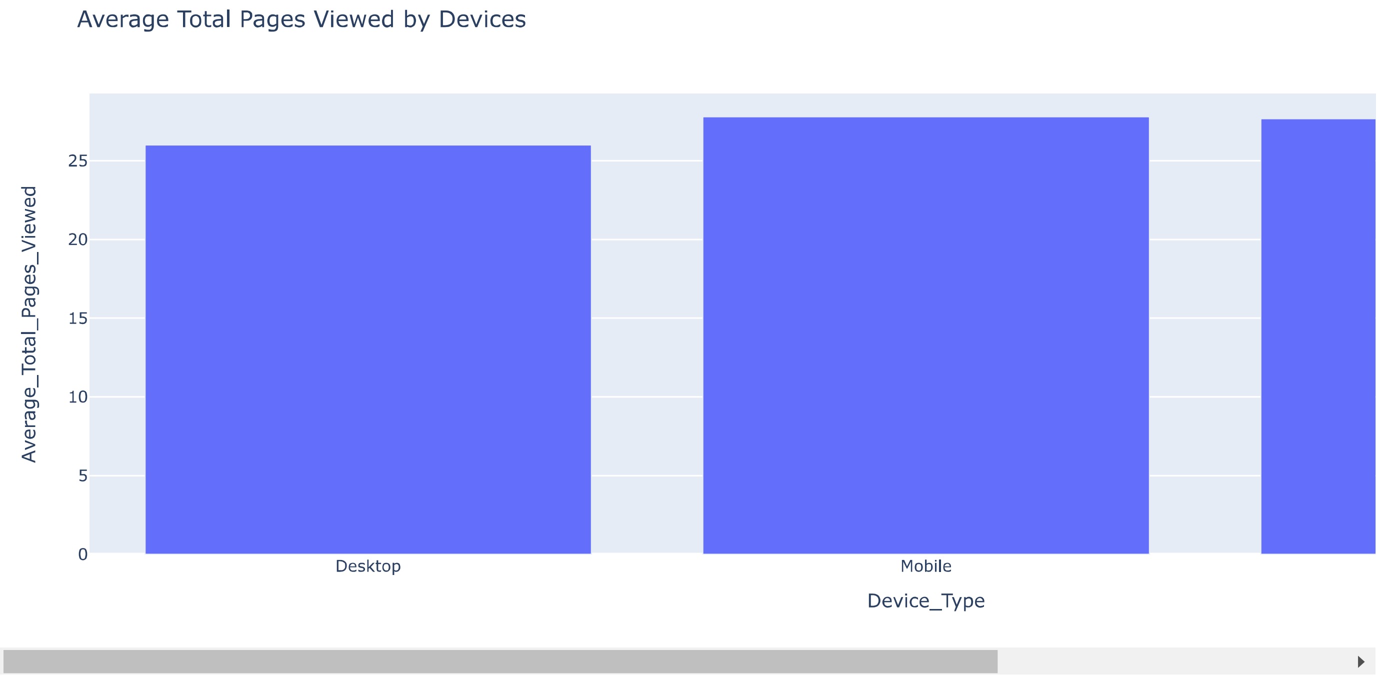
Shopping Behavior by Gender:

Females, on average, add 5.27 items to their carts, suggesting exploratory shopping. Males average 5.04 items, indicating a more focused approach. Females spend an average of $2.28 per transaction, while males spend$2.64. Understanding these gender-specific patterns enables tailored marketing strategies. **bold text**

```
# Grouped Analysis
gender_grouped = df.groupby('Gender')['Total_Pages_Viewed'].mean().reset_index()
gender_grouped.columns = ['Gender', 'Average_Total_Pages_Viewed']
fig = px.bar(gender_grouped, x='Gender', y='Average_Total_Pages_Viewed',title='Average Total Pages Viewed by Gender')
fig.show()
```



```
#Now, let’s have a look at the average total pages viewed by devices
devices_grouped = df.groupby('Device_Type')['Total_Pages_Viewed'].mean().reset_index()
devices_grouped.columns = ['Device_Type', 'Average_Total_Pages_Viewed']
fig = px.bar(devices_grouped, x='Device_Type', y='Average_Total_Pages_Viewed',
             title='Average Total Pages Viewed by Devices')
fig.show()
```



Now, let’s calculate the customer lifetime value and visualize segments based on the customer lifetime value:

```
df['CLV'] = (df['Total_Purchases'] * df['Total_Pages_Viewed']) / df['Age']
df.head()
```

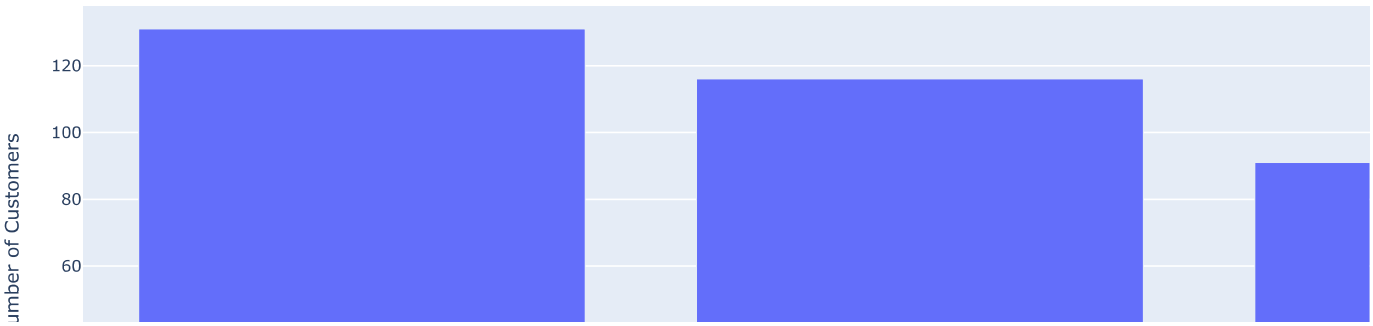
	User_ID	Gender	Age	Location	Device_Type	Product_Browsing_Time	Total_Pages_Viewed	Items_Added_to_Cart	Total_Purchases	C
0	1	Female	23	Ahmedabad	Mobile	60	30	1	0	0.0000
1	2	Male	25	Kolkata	Tablet	30	38	9	4	6.0800
2	3	Male	32	Bangalore	Desktop	37	13	5	0	0.0000
3	4	Male	35	Delhi	Mobile	7	20	10	3	1.7142
4	5	Male	27	Bangalore	Tablet	35	20	8	2	1.4814

```
df['Segment'] = pd.cut(df['CLV'], bins=[1, 2.5, 5, float('inf')],
                       labels=['Low Value', 'Medium Value', 'High Value'])

segment_counts = df['Segment'].value_counts().reset_index()
segment_counts.columns = ['Segment', 'Count']

# Create a bar chart to visualize the customer segments
fig = px.bar(segment_counts, x='Segment', y='Count',
             title='Customer Segmentation by CLV')
fig.update_xaxes(title='Segment')
fig.update_yaxes(title='Number of Customers')
fig.show()
```

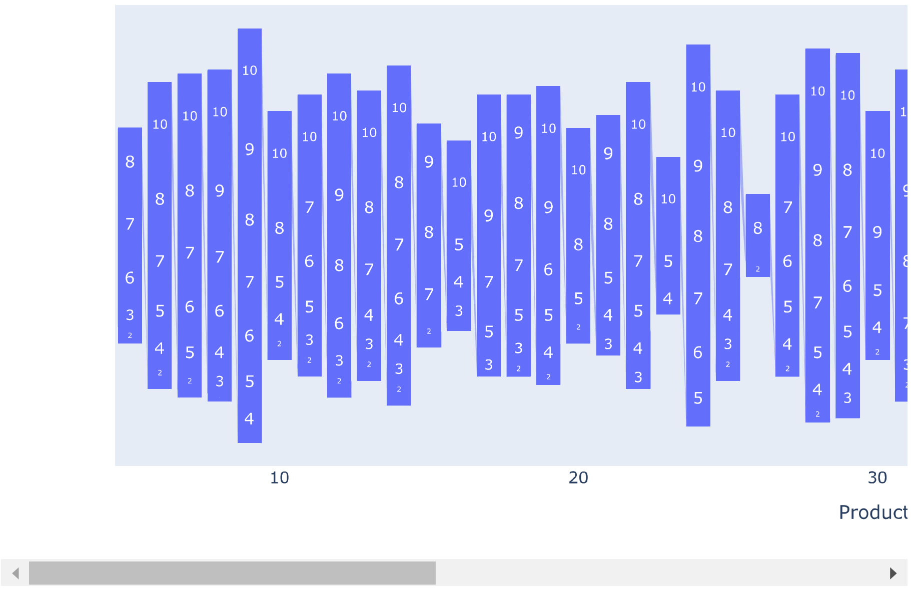
Customer Segmentation by CLV



```
# Funnel analysis
funnel_data = df[['Product_Browsing_Time', 'Items_Added_to_Cart', 'Total_Purchases']]
funnel_data = funnel_data.groupby(['Product_Browsing_Time', 'Items_Added_to_Cart']).sum().reset_index()

fig = px.funnel(funnel_data, x='Product_Browsing_Time', y='Items_Added_to_Cart', title='Conversion Funnel')
fig.show()
```

Conversion Funnel


Product