

# Customer Purchase

```
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
```

```
url = "https://github.com/Shrutinagar/datasets/raw/main/%5BPython%20Dataset%5D%20-%20cu"
```

```
from urllib.request import urlretrieve
urlretrieve(url, "customer_purchase.csv")
```

('customer\_purchase.csv', <http.client.HTTPMessage at 0x7f4dd1452dc0>)

```
df = pd.read_csv("customer_purchase.csv")
```

df

	Customer ID	Purchase Date	Product Name	Product Category	Quantity	Total Price
0	1046	2023-02-26	across	certain	17	134.96
1	1083	2023-10-10	attention	worry	13	178.36
2	1051	2023-04-04	raise	politics	13	176.99
3	1032	2023-04-16	scientist	push	6	187.37
4	1013	2023-02-26	free	word	7	103.51
...	...	...	...	...	...	...
995	1058	2022-12-02	leave	goal	9	11.15
996	1092	2023-02-03	eye	here	14	124.80
997	1007	2023-01-04	yes	over	7	12.20
998	1082	2023-08-13	toward	determine	9	97.94
999	1047	2023-11-13	worry	send	5	159.47

1000 rows × 6 columns

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Customer ID     1000 non-null  int64
1   Purchase Date   1000 non-null  object
```

```
2 Product Name      1000 non-null  object
3 Product Category  1000 non-null  object
4 Quantity          1000 non-null  int64
5 Total Price       1000 non-null  float64
```

dtypes: float64(1), int64(2), object(3)

memory usage: 47.0+ KB

## Total Revenue

```
total_revenue = round(df["Total Price"].sum(),2)
print("Total Revenue is Rs.{}".format(total_revenue))
```

Total Revenue is Rs.107285.83

## Top 5 Best Selling Products

```
best_seller = df.groupby("Product Name")["Quantity"].sum()
```

```
best_seller = best_seller.sort_values(by = "Quantity", ascending = False)
```

```
best_seller_5 = best_seller.head(5)
best_seller_5
```

	Quantity
Product Name	
special	76
degree	64
least	60
accept	55
true	54

```
plt.pie(data = df, )
```

## Average Quantity of Products Purchased per Customer

```
Avg_Quantity = round(df.groupby("Customer ID")["Quantity"].mean(),0)
```

```
with pd.option_context("display.max_rows",None):
    print(Avg_Quantity)
```

Customer ID

1001      9.0

1002	10.0
1003	8.0
1004	8.0
1005	10.0
1006	9.0
1007	10.0
1008	9.0
1009	12.0
1010	11.0
1011	11.0
1012	12.0
1013	10.0
1014	6.0
1015	10.0
1016	8.0
1017	14.0
1018	10.0
1019	9.0
1020	12.0
1021	10.0
1022	11.0
1023	10.0
1024	9.0
1025	15.0
1026	11.0
1027	10.0
1028	10.0
1029	13.0
1030	13.0
1031	9.0
1032	10.0
1033	8.0
1034	8.0
1035	13.0
1036	9.0
1037	10.0
1038	10.0
1039	7.0
1040	10.0
1041	10.0
1042	6.0
1043	10.0
1044	11.0

1045	10.0
1046	9.0
1047	9.0
1048	16.0
1049	9.0
1050	11.0
1051	8.0
1052	9.0
1053	6.0
1054	10.0
1055	12.0
1056	12.0
1057	10.0
1058	11.0
1059	10.0
1060	12.0
1061	11.0
1062	13.0
1063	14.0
1064	9.0
1065	8.0
1066	12.0
1067	9.0
1068	15.0
1069	12.0
1070	9.0
1071	9.0
1072	9.0
1073	9.0
1074	14.0
1075	8.0
1076	15.0
1077	12.0
1078	14.0
1079	9.0
1080	10.0
1081	12.0
1082	13.0
1083	11.0
1084	9.0
1085	10.0
1086	11.0
1087	10.0

1088	12.0
1089	10.0
1090	8.0
1091	10.0
1092	11.0
1093	9.0
1094	11.0
1095	12.0
1096	9.0
1097	11.0
1098	10.0
1099	7.0
1100	6.0

Name: Quantity, dtype: float64

## Monthly Sales Report

-Total Revenue for each month

```
# converting Purchase Date column to date format
df['Purchase Date'] = pd.to_datetime(df["Purchase Date"])

# Creating new column for month
df["Month"] = df["Purchase Date"].dt.to_period("M")
```

```
df["Year"] = df["Purchase Date"].dt.year
```

```
#monthly sales report
monthly_sales = df.groupby("Month")["Total Price"].sum()
```

```
print(monthly_sales)
```

Month	
2022-11	578.79
2022-12	9414.71
2023-01	9451.95
2023-02	9634.90
2023-03	8997.52
2023-04	9049.23
2023-05	6178.53
2023-06	8664.21
2023-07	10132.95
2023-08	7860.04
2023-09	8615.01

2023-10      7933.93

2023-11      10774.06

Freq: M, Name: Total Price, dtype: float64