03bikebuyers

August 1, 2025

1 Bike buyers

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

[2]:
```
df = pd.read_csv("../../datasets/bike_buyers_clean.csv")
```

[3]:
```
df.head()
```

[3]:

	ID	Marital Status	Gender	Income	Children	Education	Occupation	Home Owner	Cars	Commute Distance	Region	Age	Purchased Bike
0	12496	Married	Female	40000	1	Bachelors	Skilled Manual	Yes	0	0-1 Miles	Europe	42	No
1	24107	Married	Male	30000	3	Partial College	Clerical	Yes	1	0-1 Miles	Europe	43	No
2	14177	Married	Male	80000	5	Partial College	Professional	No	2	2-5 Miles	Europe	60	No
3	24381	Single	Male	70000	0	Bachelors	Professional	Yes	1	5-10 Miles	Pacific	41	Yes
4	25597	Single	Male	30000	0	Bachelors	Clerical	No	0	0-1 Miles	Europe	36	Yes

1.1 Što nas interesira

- Bračni status osoba sa motociklima, želimo pokazati da samci više voze motore
- Vidjeti po spolu, obrazovanje i regija tko više kupuje motocikle
  - Pretpostavka je da obazovaniji ljudi manje voze motocikle
- Vidjeti koliko osoba ima djece - pretpostavka je da ljudi sa djecom ne žele riskirati
- Vidjeti koliko osoba ima automobila - pretpostavka je da ljudi sa više automobila ne žele plaćati dodatna osgiranja i registracije

[4]:
```
df_bracni_status = df.groupby("Marital Status")
```

[5]:
```
df_bracni_status["Marital Status"].value_counts()
```

[5]:
```
Marital Status
Married    539
Single     461
Name: count, dtype: int64
```

[6]:
```
married = df_bracni_status["Marital Status"].value_counts()["Married"]
single = df_bracni_status["Marital Status"].value_counts()["Single"]
print(married/(married + single)*100)
```

```
53.900000000000006
```

[7]:
```
df_bracni_status["Purchased Bike"].value_counts()
```

[7]:
```
Marital Status  Purchased Bike
Married         No               307
                Yes              232
Single          Yes              249
                No               212
Name: count, dtype: int64
```

Zaključak: Samci kupuju više, ali ne značajno.

[8]:
```
df_sex = df.groupby("Gender")
df_edu = df.groupby("Education")
df_region = df.groupby("Region")
```

[9]:
```
df_sex["Purchased Bike"].value_counts()
```

[9]:
```
Gender  Purchased Bike
Female  No               252
        Yes              239
Male    No               267
        Yes              242
Name: count, dtype: int64
```

[10]:
```
df_edu["Purchased Bike"].value_counts()
```

[10]:
```
Education            Purchased Bike
Bachelors            Yes              169
                     No               137
Graduate Degree      Yes               94
                     No                80
High School          No               100
                     Yes               79
Partial College      No               146
```

```
                      Yes               119
Partial High School   No                 56
                      Yes                20
Name: count, dtype: int64
```

[11]:
```
df_region["Purchased Bike"].value_counts()
```

[11]:
```
Region         Purchased Bike
Europe         No                152
               Yes               148
North America  No                288
               Yes               220
Pacific        Yes               113
               No                 79
Name: count, dtype: int64
```

Zaključak: Neznatno više muškaraca kupuje motocikle, obrazovani više kupuju motocikle, a najviše motocikala je prodano u North America.

[12]:
```
df_children = df.groupby(["Children"])
df_children["Purchased Bike"].value_counts()
```

[12]:
```
Children  Purchased Bike
0         Yes               140
          No                137
1         Yes                98
          No                 72
2         No                112
          Yes                98
3         Yes                73
          No                 62
4         No                 73
          Yes                54
5         No                 63
          Yes                18
Name: count, dtype: int64
```

Zaključak: Ova teza stoji, ljudi sa više djece manje kupuju motocikle.

[13]:
```
df_cars = df.groupby(["Cars"])
df_cars["Purchased Bike"].value_counts()
```

[13]:
```
Cars  Purchased Bike
0     Yes               150
      No                 93
1     Yes               152
      No                115
2     No                220
```

```
      Yes              125
3     No                52
      Yes               33
4     No                39
      Yes               21
Name: count, dtype: int64
```

Zaključak: Ova teza stoji, ljudi sa više automobila manje kupuju motocikle.

[14]:
```
df_two = df.groupby(["Purchased Bike", "Gender"])
```

[15]:
```
df_two["Income"].agg(["count", "mean", "min", "max"])
```

[15]:

		count	mean	min	max
Purchased Bike	Gender				
No	Female	252	53650.793651	10000	130000
	Male	267	56029.962547	10000	170000
Yes	Female	239	55648.535565	10000	170000
	Male	242	59338.842975	10000	160000

[16]:
```
df_two["Age"].agg(["count", "mean", "min", "max"])
```

[16]:

		count	mean	min	max
Purchased Bike	Gender				
No	Female	252	45.444444	25	89
	Male	267	45.258427	26	73
Yes	Female	239	42.736402	25	72
	Male	242	43.140496	25	78

[17]:
```
df_two["Home Owner"].value_counts()
```

[17]:
```
Purchased Bike  Gender  Home Owner
No              Female  Yes           171
                        No             81
                Male    Yes           189
                        No             78
Yes             Female  Yes           163
                        No             76
                Male    Yes           162
                        No             80
Name: count, dtype: int64
```

[18]:
```
df_3 = df.groupby(["Purchased Bike", "Gender", "Home Owner"])
```

[19]:
```
df_3["Income"].agg(["count", "mean", "min", "max"])
```

[19]:

			count	mean	min	max
Purchased Bike	Gender	Home Owner				

No	Female	No	81	50740.740741	10000	130000
		Yes	171	55029.239766	10000	130000
	Male	No	78	54102.564103	10000	170000
		Yes	189	56825.396825	10000	170000
Yes	Female	No	76	52368.421053	10000	170000
		Yes	163	57177.914110	10000	150000
	Male	No	80	64000.000000	10000	160000
		Yes	162	57037.037037	10000	160000

Zaključak: Oni koji kupuju već imaju kuću i to su podjenako muškarci i žene.