Ninjaz

STC
الاتصالات السعودية

CAPSTONE PROJECT

PRESENTATION OUTLINE

- STC Dataset
- STC and Saudi vision 2030
- Exploratory Data Analysis (EDA)
- Data cleaning
- Data pre-processing

STC DATASET

```
RangeIndex: 1048575 entries, 0 to 1048574
Data columns (total 20 columns):
 #   Column            Non-Null Count    Dtype
---  ------            --------------    -----
 0   CAL_DT            1048575 non-null  object
 1   MODEL_NAME        1048575 non-null  object
 2   BRAND_FULL_NAME   1048575 non-null  object
 3   BRAND_NAME        1048575 non-null  object
 4   VENDOR_NAME       1048575 non-null  object
 5   OS_NAME           1048575 non-null  object
 6   DEVICE_TYPE       1048575 non-null  object
 7   _2G_FLG           1048575 non-null  object
 8   _3G_FLG           1048575 non-null  object
 9   _4G_FLG           1048575 non-null  object
 10  WIFI_FLG          1048575 non-null  object
 11  BLUETOOTH_FLG     1048575 non-null  object
 12  TOUCH_SCREEN_FLG  1048575 non-null  object
 13  DUAL_SIM_FLG      1048575 non-null  object
 14  GENDER_TYPE_CD    939245 non-null   object
 15  AGE_B             1048575 non-null  object
 16  NATIONALITY_CD    925709 non-null   object
 17  NATIONALITY_NAME  925933 non-null   object
 18  SAUDI_NON_SAUDI   1048082 non-null  object
 19  DEVICE_COUNT      1048086 non-null  object
dtypes: object(20)
memory usage: 160.0+ MB
```

The data set describes uncommon handset devices usage by customers, for an interval of 12 months and with specific customer demographics. It can be used to analyze some devices trends over time, and the devices used by different groups of customers.

The dataset contains 714023 rows, and has the following attributes:

stc AND SAUDI VISION 2030

5

Vital Society

strengthening the economy of Saudi Arabia

Thriving Economy

STC launched the Saudi Vision Cable project

Ambitious Nation

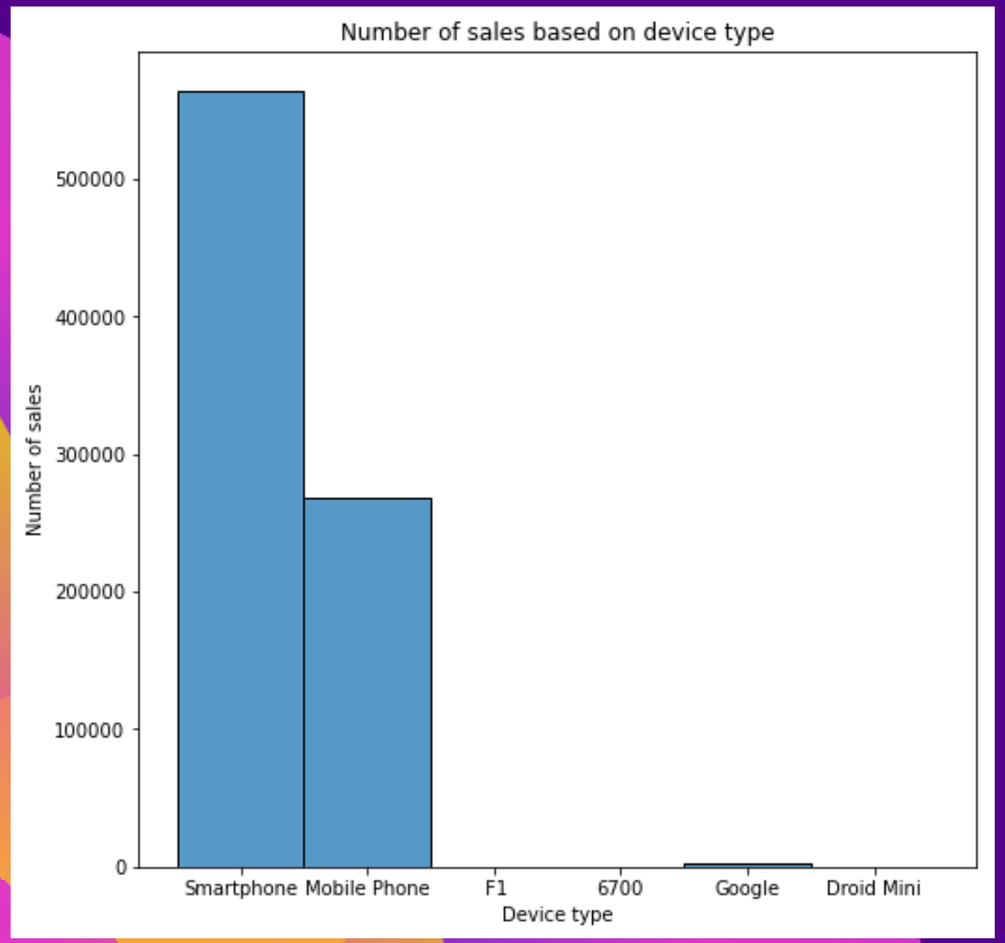
empower the Saudis and the private companies to take better steps and continue improving

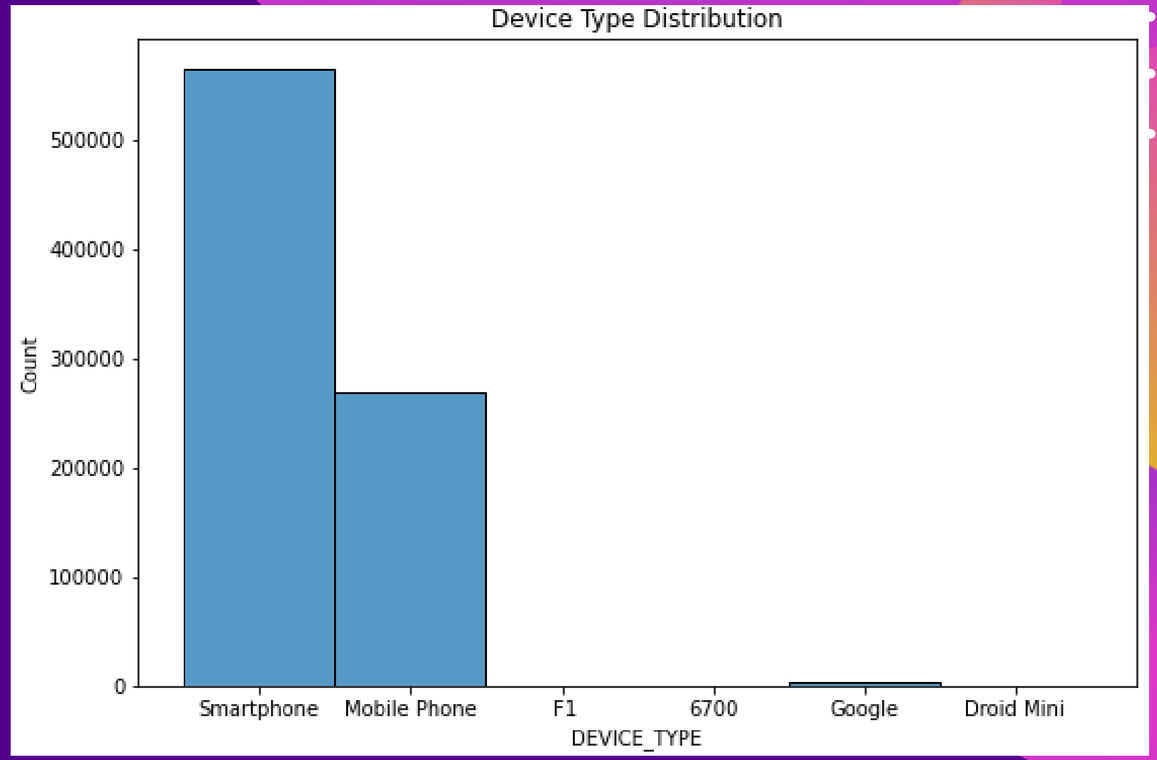
EDA

03

Sales

Number of sales based on device type

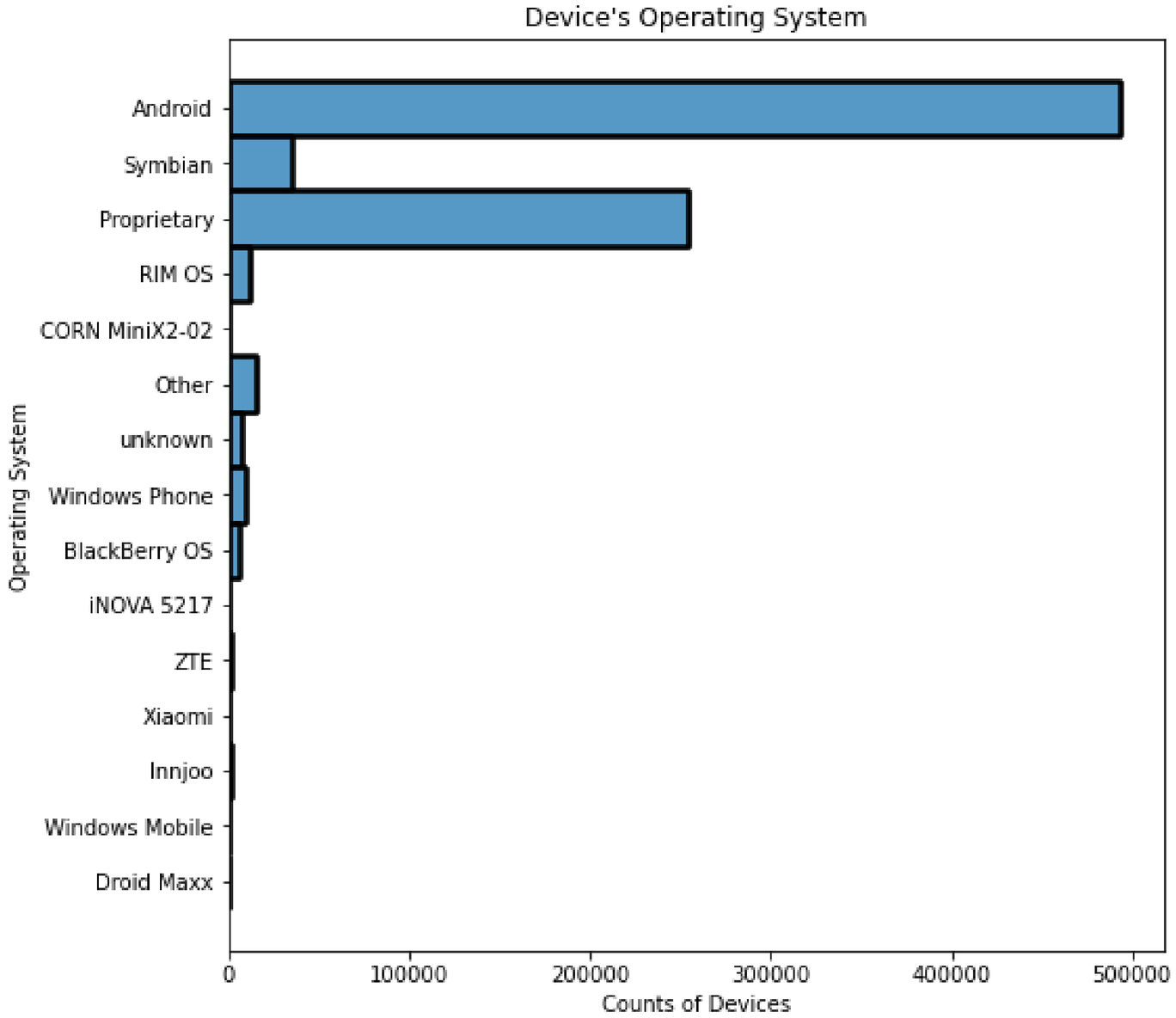




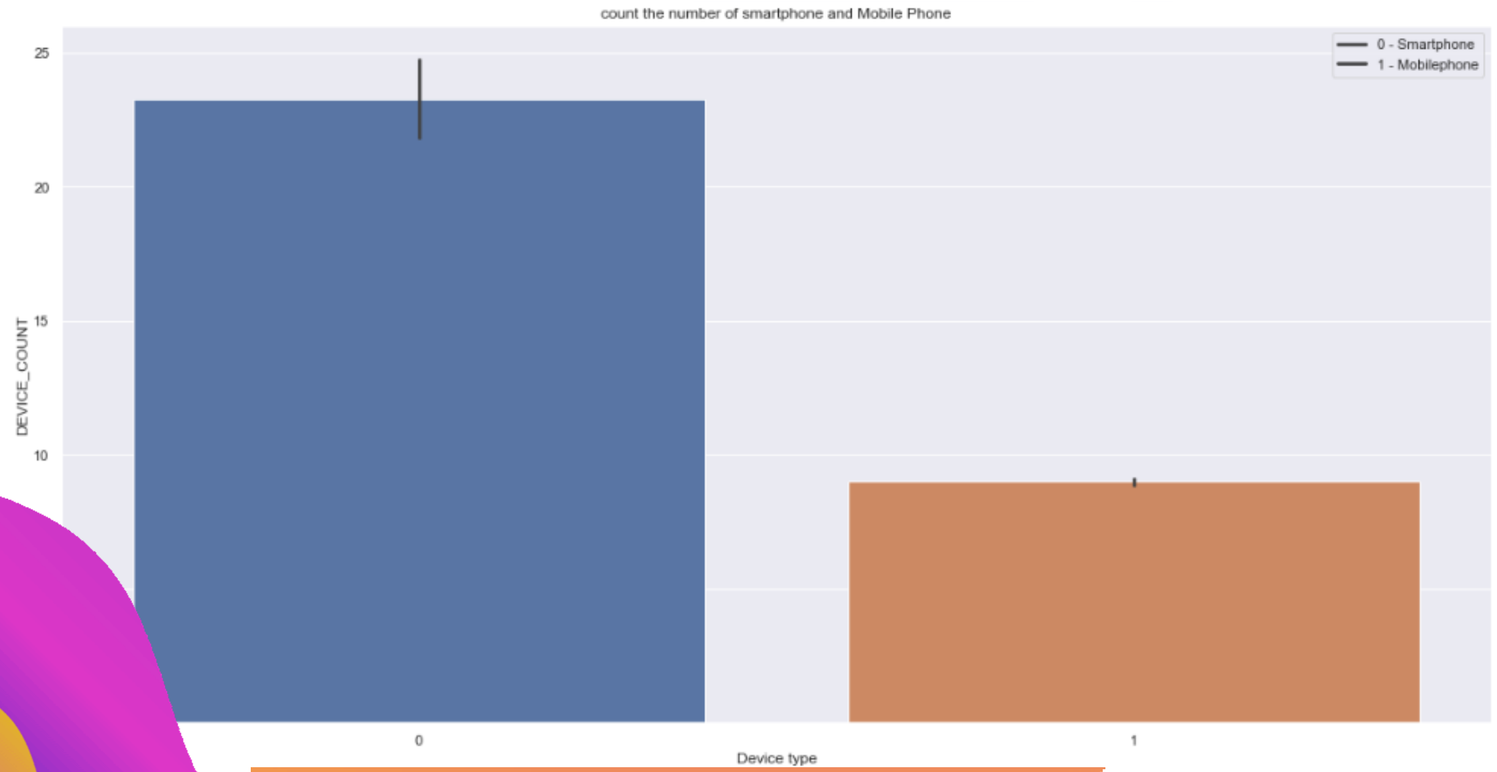
Distribution

Number of devices sold based on type

OPERATING SYSTEMS



This is the count of devices sold based on operating system

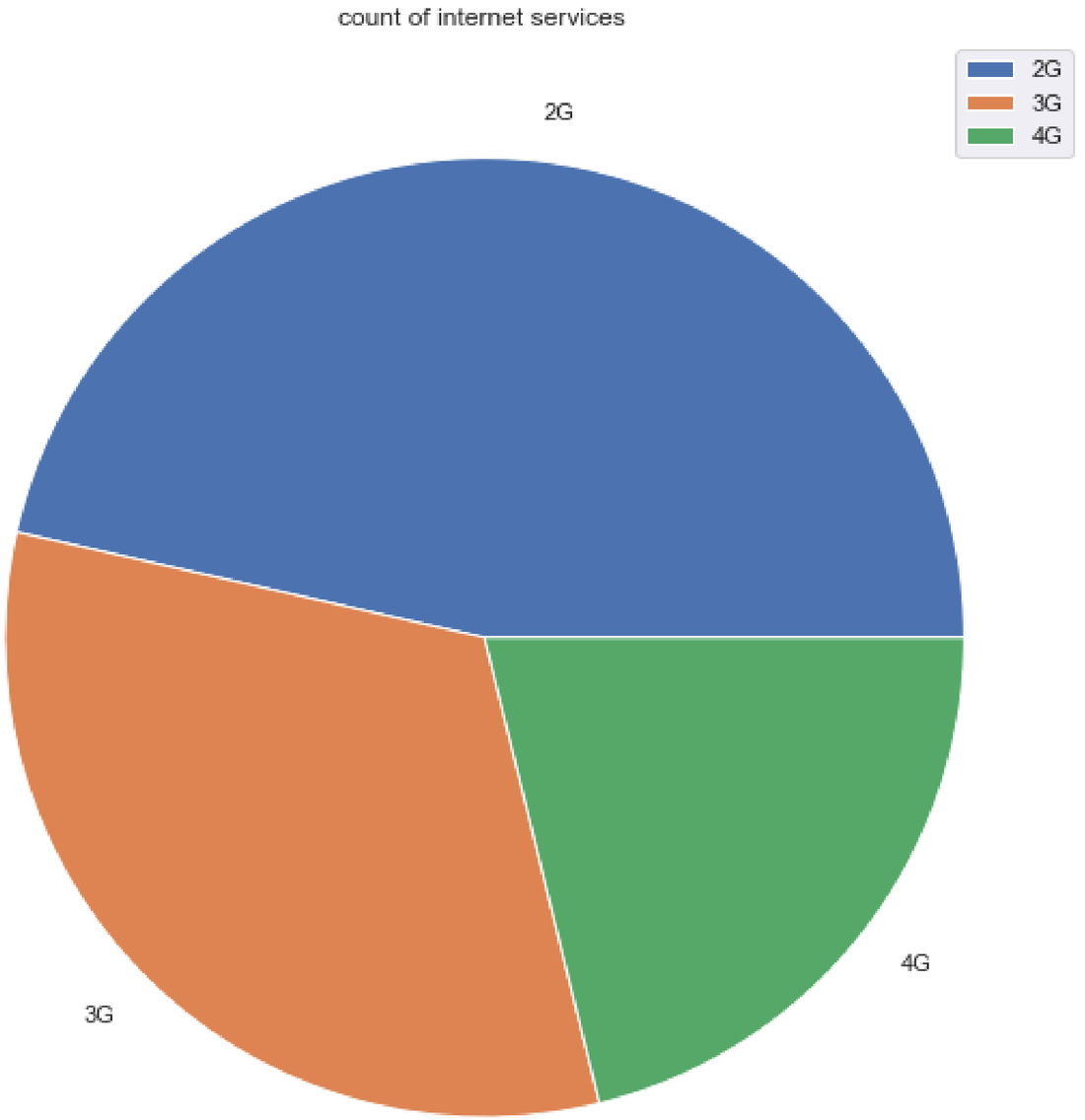


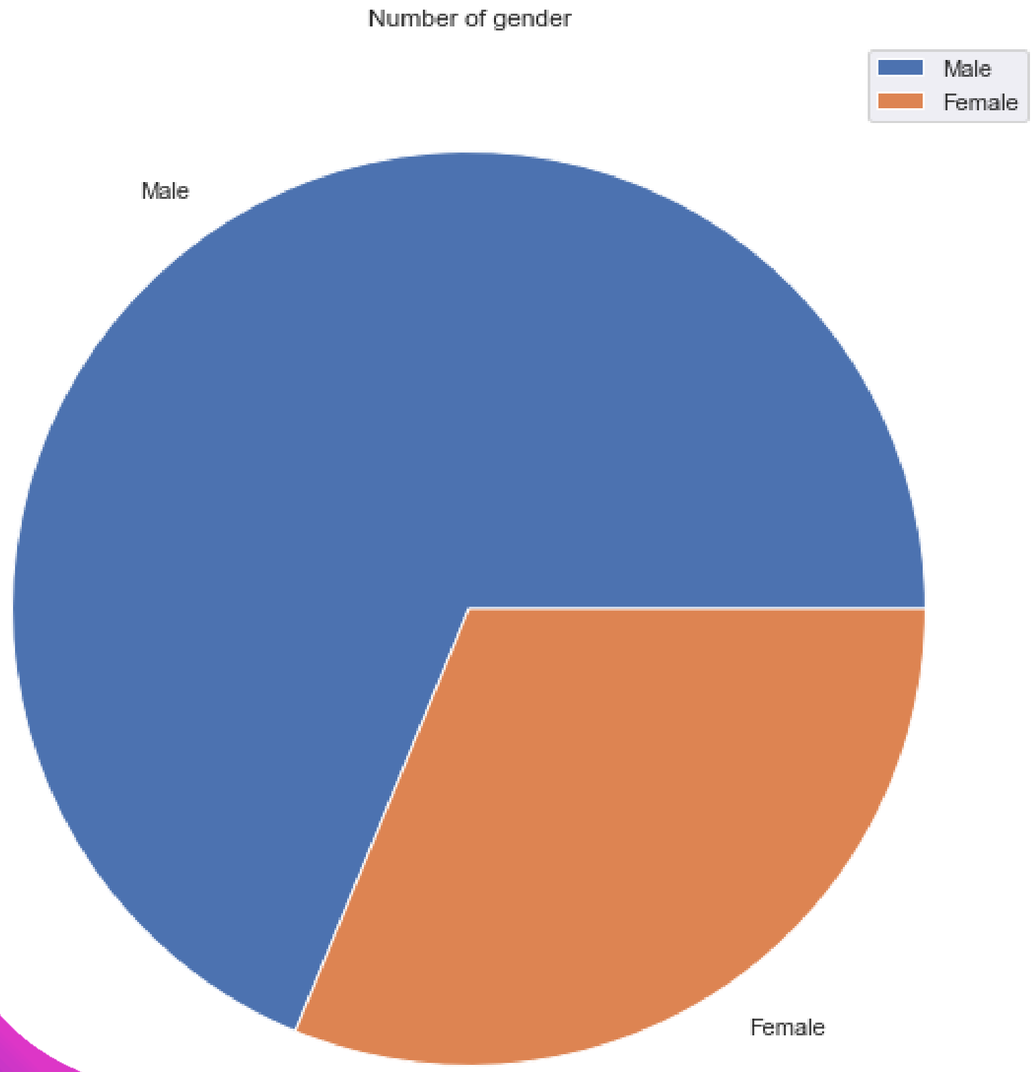
Smartphone Mobile Phone

this chart shows the count of smartphones and mobile phones devices

INTERNET SERVICES

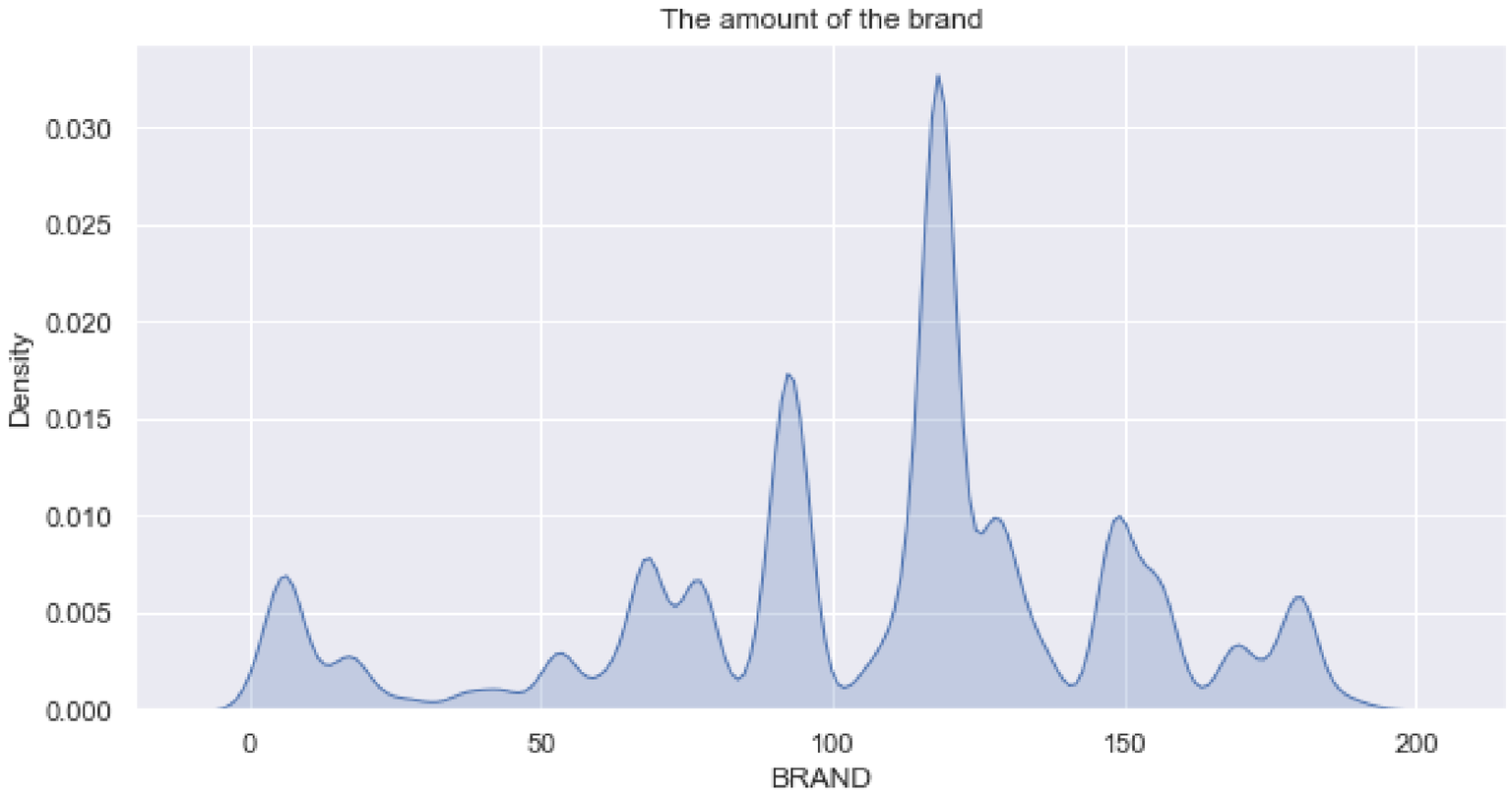
the count of devices that have 2G, 3G and 4G internet service





Sales Based on Gender

this chart shows number of devices sold based on gender

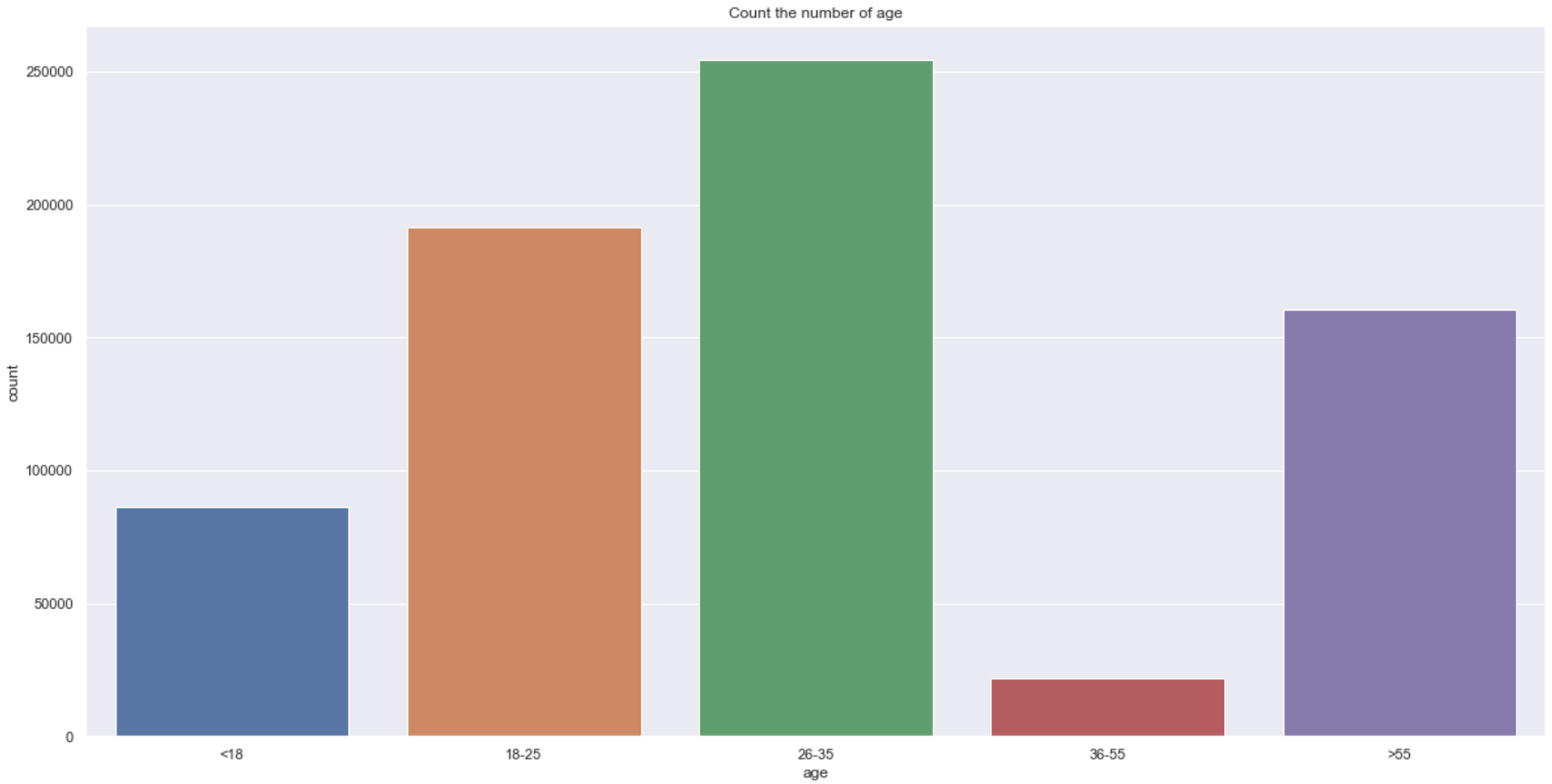


The number of the brand

this chart shows the number of each brand

THE NUMBER OF AGE

the count number of each age



DATA CLEANING

step 1

finding unique values for all features

```
stc_d['AGE_B'].unique()
```

step 2

dropping unnecessary values

```
stc_d.drop(stc_d[(stc_d['NATIONALITY_NAME']=='18-25') | (stc_d['NATIONALITY_NAME']== '<18  ') |
                 (stc_d['NATIONALITY_NAME']== 'N') | (stc_d['NATIONALITY_NAME']== 'NA   ')].index, axis=0,inplace=True)
```

step 3

removing unnecessary cilumns

```
stc_d.drop(columns="BRAND_FULL_NAME",axis=1,inplace=True)
```

step 4

removing duplicated columns after encoding

```
le = preprocessing.LabelEncoder()
stc_d["2G_FLG"]=le.fit_transform(stc_d["_2G_FLG"])
stc_d["3G_FLG"]=le.fit_transform(stc_d["_3G_FLG"])
stc_d["4G_FLG"]=le.fit_transform(stc_d["_4G_FLG"])
stc_d["WIFI"]=le.fit_transform(stc_d["WIFI_FLG"])
stc_d["BLUETOOTH"]=le.fit_transform(stc_d["BLUETOOTH_FLG"])
stc_d["TOUCH_SCREEN"]=le.fit_transform(stc_d["TOUCH_SCREEN_FLG"])
stc_d["DUAL_SIM"]=le.fit_transform(stc_d["DUAL_SIM_FLG"])
stc_d["GENDER"]=le.fit_transform(stc_d["GENDER_TYPE_CD"])
stc_d["MODEL"]=le.fit_transform(stc_d["MODEL_NAME"])
stc_d["BRAND"]=le.fit_transform(stc_d["BRAND_NAME"])
stc_d["VENDOR"]=le.fit_transform(stc_d["VENDOR_NAME"])
stc_d["OS"]=le.fit_transform(stc_d["OS_NAME"])
stc_d["DEVICE"]=le.fit_transform(stc_d["DEVICE_TYPE"])
stc_d["AGE"]=le.fit_transform(stc_d["AGE_B"])
stc_d["NATIONALITY"]=le.fit_transform(stc_d["NATIONALITY_NAME"])
stc_d["SAUDI"]=le.fit_transform(stc_d["SAUDI_NON_SAUDI"])
```

DATA PREPROCESING

step 2

Using label encoder on the columns

```
le = preprocessing.LabelEncoder()
stc_d["2G_FLG"]=le.fit_transform(stc_d["_2G_FLG"])
stc_d["3G_FLG"]=le.fit_transform(stc_d["_3G_FLG"])
stc_d["4G_FLG"]=le.fit_transform(stc_d["_4G_FLG"])
stc_d["WIFI"]=le.fit_transform(stc_d["WIFI_FLG"])
stc_d["BLUETOOTH"]=le.fit_transform(stc_d["BLUETOOTH_FLG"])
stc_d["TOUCH_SCREEN"]=le.fit_transform(stc_d["TOUCH_SCREEN_FLG"])
stc_d["DUAL_SIM"]=le.fit_transform(stc_d["DUAL_SIM_FLG"])
stc_d["GENDER"]=le.fit_transform(stc_d["GENDER_TYPE_CD"])
stc_d["MODEL"]=le.fit_transform(stc_d["MODEL_NAME"])
stc_d["BRAND"]=le.fit_transform(stc_d["BRAND_NAME"])
stc_d["VENDOR"]=le.fit_transform(stc_d["VENDOR_NAME"])
stc_d["OS"]=le.fit_transform(stc_d["OS_NAME"])
stc_d["DEVICE"]=le.fit_transform(stc_d["DEVICE_TYPE"])
stc_d["AGE"]=le.fit_transform(stc_d["AGE_B"])
stc_d["NATIONALITY"]=le.fit_transform(stc_d["NATIONALITY_NAME"])
stc_d["SAUDI"]=le.fit_transform(stc_d["SAUDI_NON_SAUDI"])
```

step 1

Changing column types

```
stc_d["CAL_DT"]=pd.to_datetime(stc_d["CAL_DT"])
```

```
stc_d["CAL_DT"]=pd.to_datetime(stc_d["CAL_DT"]).dt.strftime('%Y')
```

```
stc_d["DEVICE_COUNT"]=stc_d["DEVICE_COUNT"].astype(str).astype(int)
```

References

- **STC** https://www.stc.com.sa/
- **Vision 2030** https://www.vision2030.gov.sa
- **STC**: The Change Management Process and the Saudi 2030 vision. | LinkedIn

THANK YOU