

TREND ANALYSIS WITH PHONE DATA

TOOLS USED

➤ Python

➤ Jupyter Notebook

➤ Libraries used:

- Numpy

- Sklearn

- Pandas

- mpl_toolkits

- Seaborn

- Matplotlib

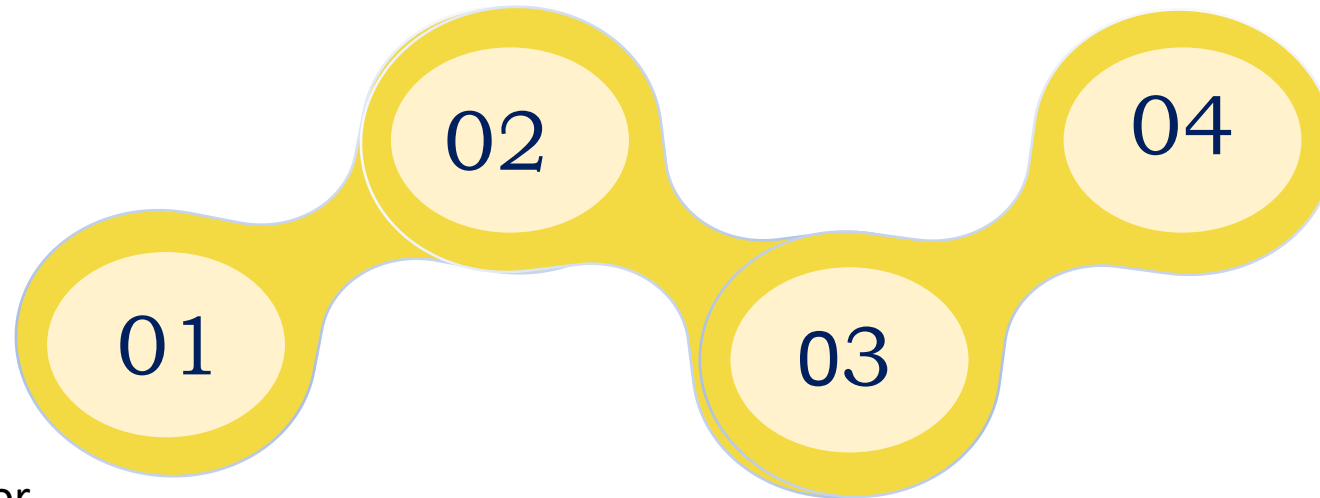
STEPS INVOLVED

Understanding Dataset and its Integrity :

What the dataset contains and finding the goodness of the data source

Data Visualization :

Using various representations in python for analysing columns in the dataset



Loading Dataset:

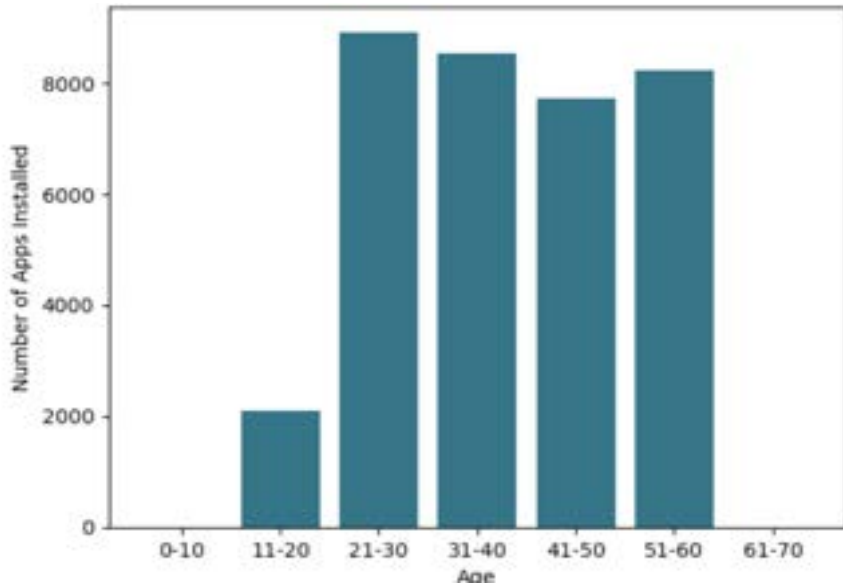
To bring our dataset to the code environment for further analysis

Statistical Analysis :

Using statistical functions like mean, median, range, standard deviation, etc,.

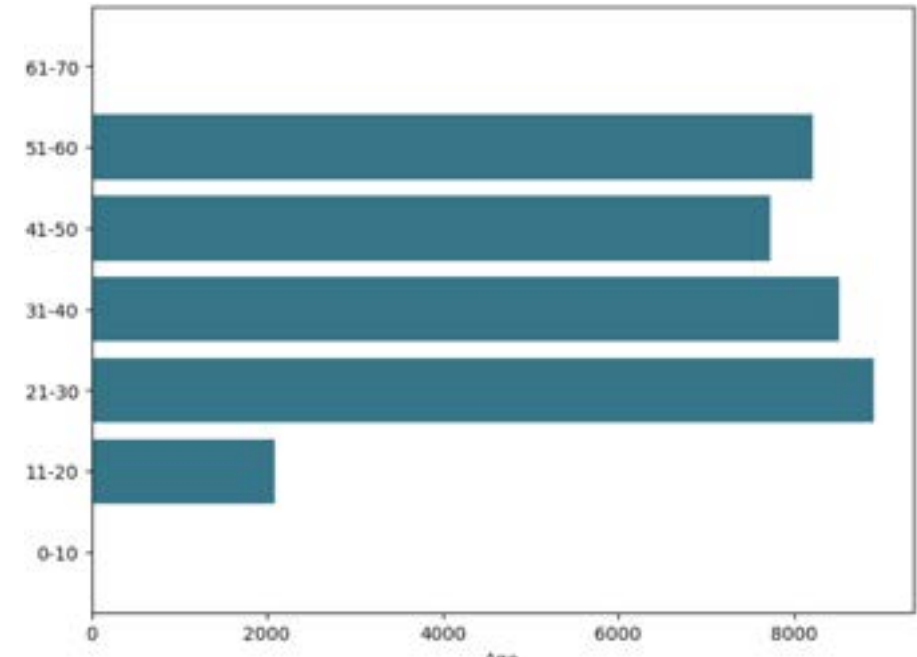
DATA VISUALIZATION

BAR CHART



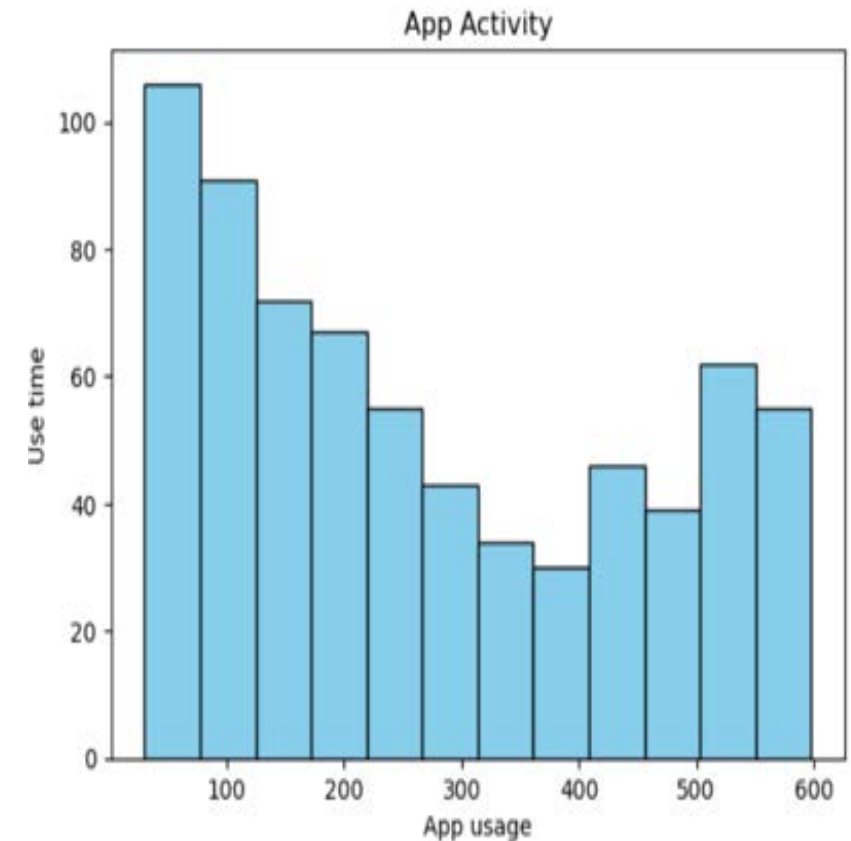
- **Represents** the relation between different age groups and the sum of the respective age groups apps installed.
- **Function** used to create : `bar()`
- **Showcases** that, most of the age groups have a similar sum number of installed apps

- **Concludes** that the age groups 21–30 and 31–40 have the highest total app installations, indicating that young adults use the most apps.
- **Function** used to create : `barh()`



HISTOGRAM

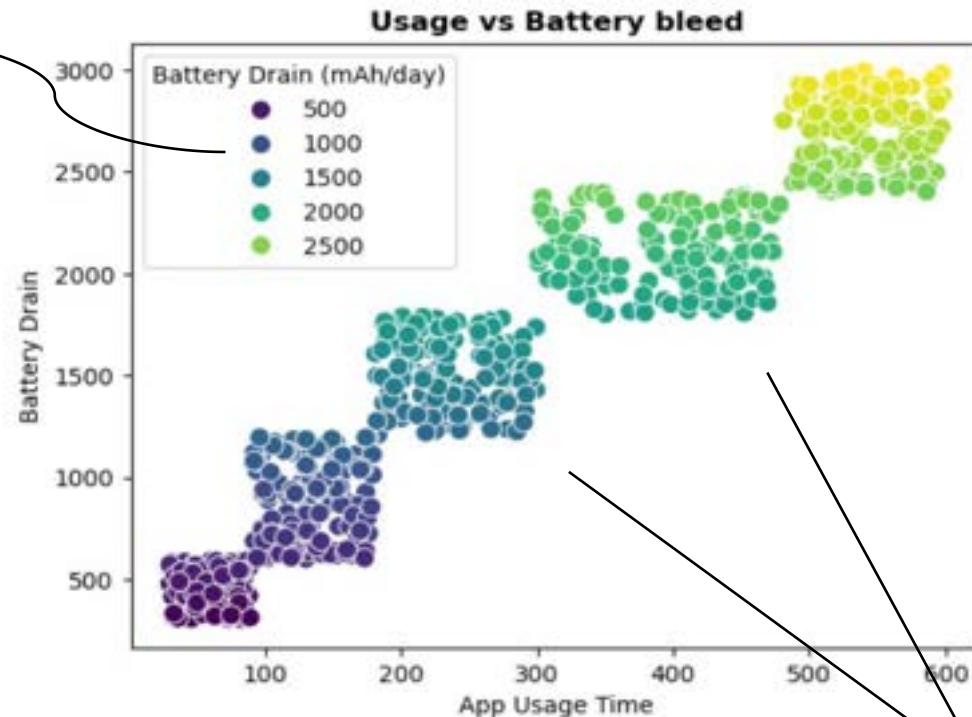
- **Represents** how the app usage time is elongated with respect to number of data points
- X-axis represent time intervals (in minutes per day).
- Y-axis indicating the frequency of days within each interval.
- **Function** used to create : `hist()`
- Shows that the app usage times is a **right-skewed distribution**.
- **Reveals** that most users tend to spend typical amount of time using apps each day.
- Higher frequencies in certain usage time ranges and fewer days with extremely low or high usage.



SCATTER PLOT

- **Represents** the relationship between app usage time (in minutes per day) and battery drain (in mAh/day).

Shows colours of the points associated with Battery Drain



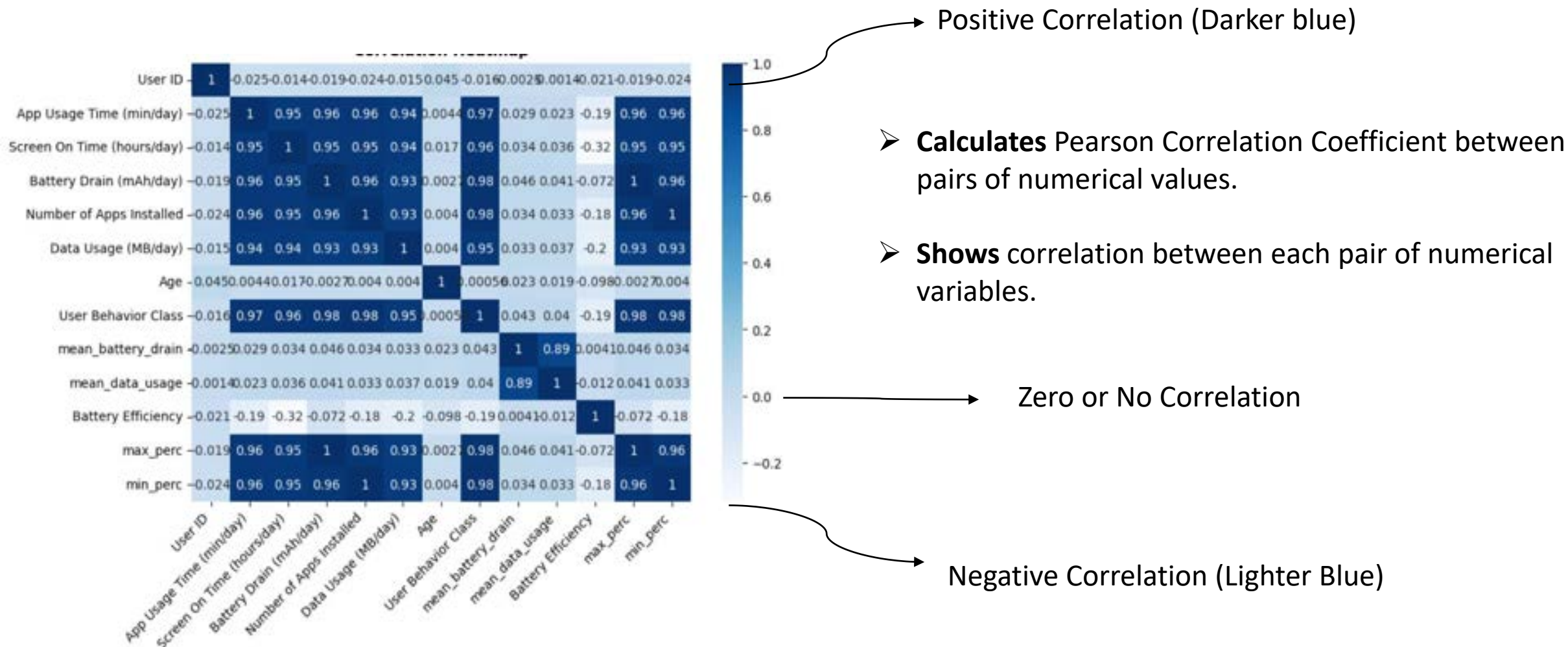
Yellow points highlights **highest battery drain** value

Suggests that users who use apps more tend to have higher battery drain.

Positive Correlation as points rise from left to right

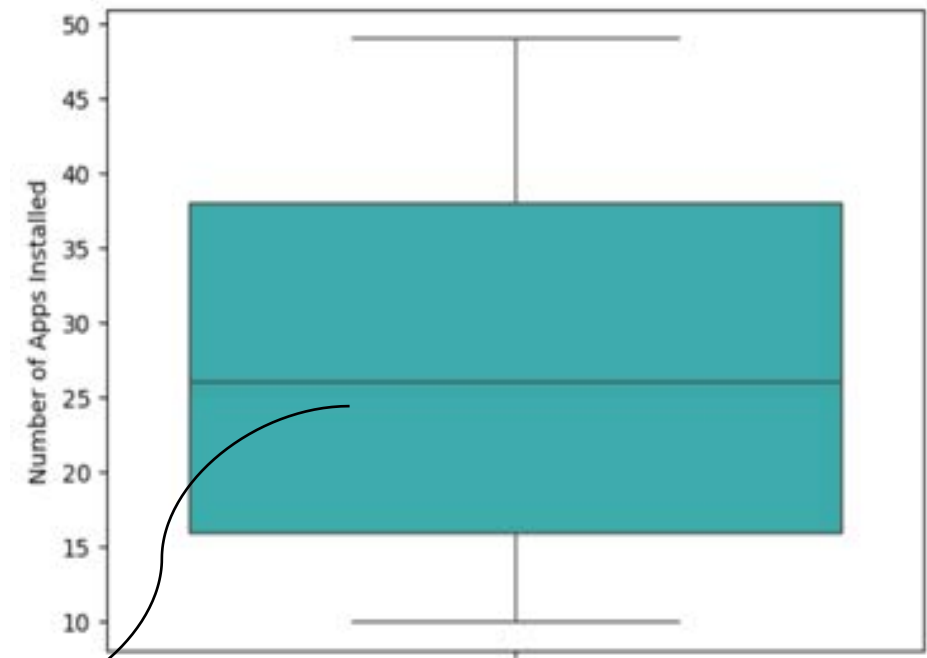
Forming Clusters which shows distinct user behaviour.

CORRELATION HEATMAP

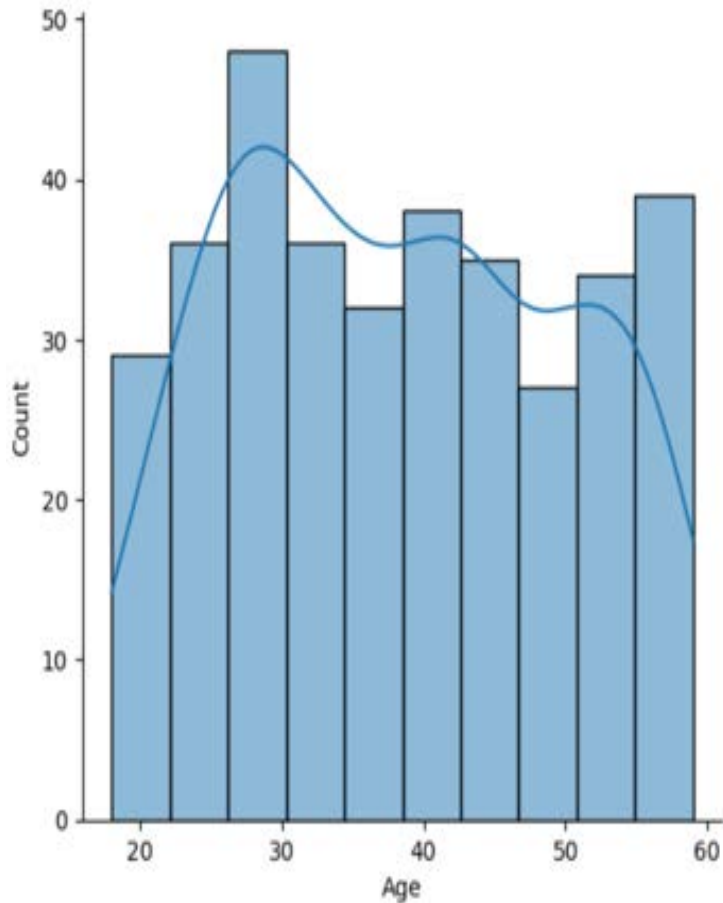


BOX PLOT

- **Provides** visual summary of number of apps installed, highlighting the median, quartiles and outliers.
- The whiskers extend from the box to min and max values within 1.5 times the interquartile range(IQR).
- **Benefits** in finding outliers that fall outside the whiskers and are as individual points or dots outside whiskers.
 - The median (50th percentile) gives robust measure of central value, which is less affected by outliers compared to the mean



DISTRIBUTION PLOT WITH KDE



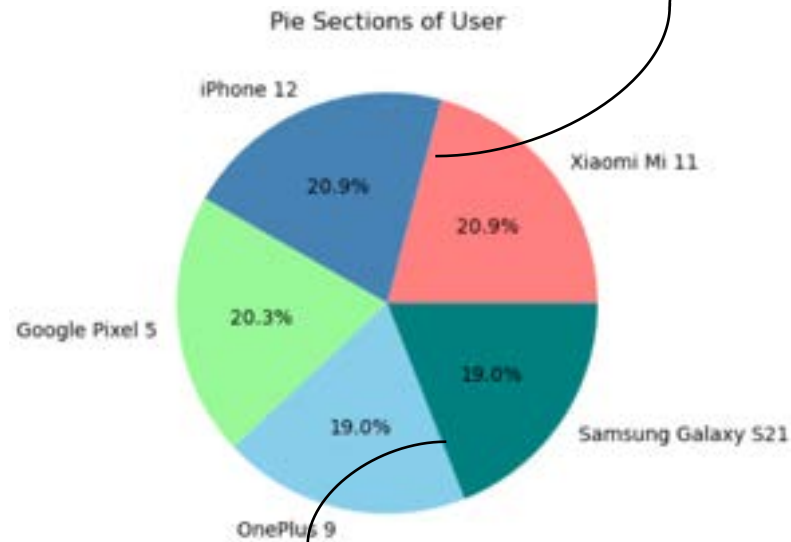
- **Generates** a plot that shows the distribution of ages in the dataset.
- **Shows** a smoothed, continuous curve that estimates the probability density function of the age distribution
- This histogram will display frequency of different age ranges.
- **KDE curve** will overlay a smooth line that estimates the density of the age distribution.
- shows a smoothed, continuous curve that estimates the probability density function of the age distribution
 - We can **conclude** that the dataset is bimodal distribution with peaks in the late 20s early 30s and late 50s, indicating two main age groups with higher concentrations of users.

PIE CHART

➤ **Represents** the distribution of users across five mobile device models.

➤ **Reveals** relatively an even distribution among the devices with :

○ Xiaomi Mi and iPhone 12 each hold the largest share at 20.9%.



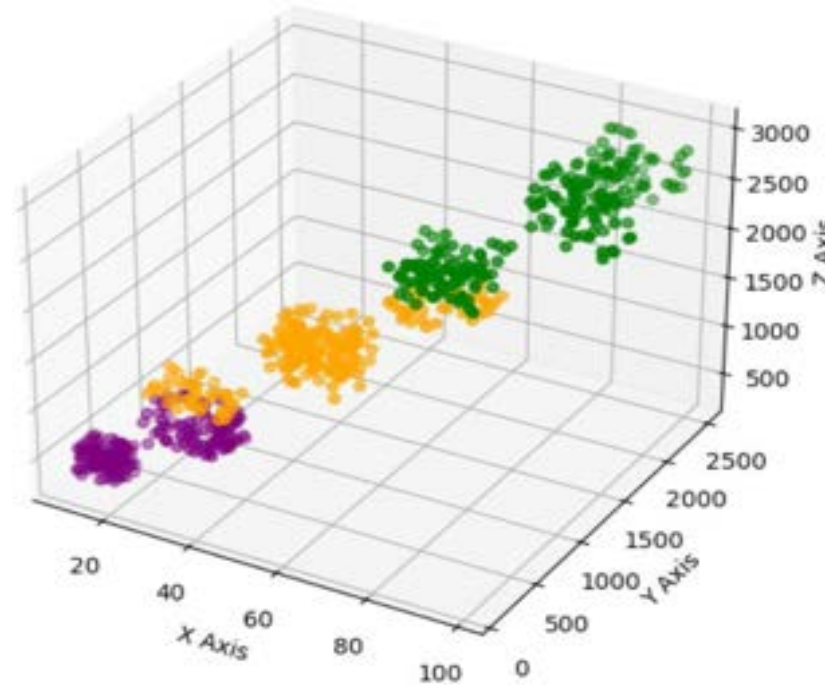
➤ The labels on the pie chart indicate the different device models.

➤ The size of each slice corresponds to the number of users for that device model.

○ OnePlus 9 and Samsung Galaxy S21 at 19.0%

3D SCATTER PLOT

Reveals a clear upward trend, showing that users with more installed apps tend to have higher data usage and greater battery drain



Purple : represents low battery drain
(<1000 mAh/day)

Green : represents high battery drain
(>2000 mAh/day)

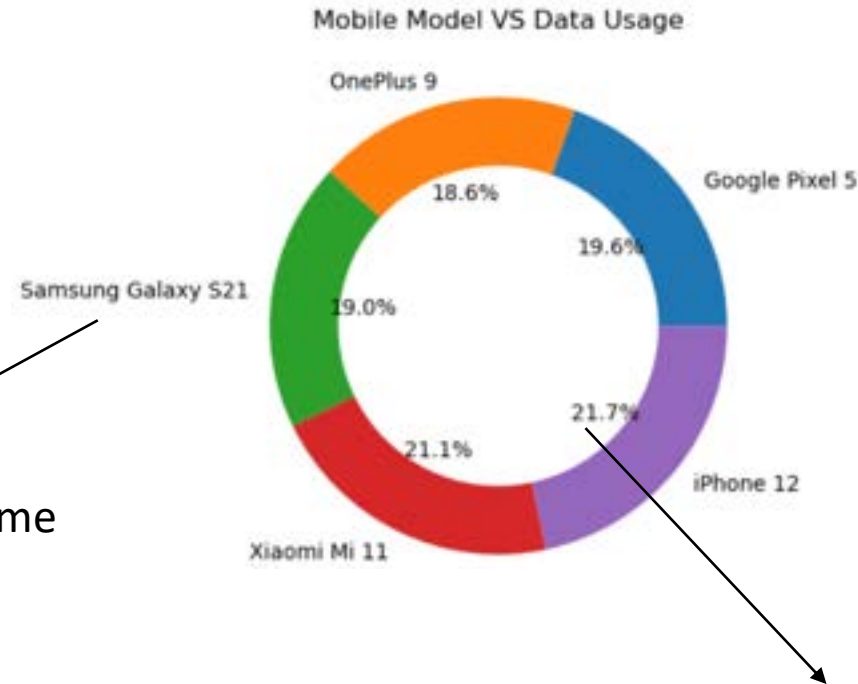
Orange : represents moderate battery drain
(1000 - 2000 mAh/day)

DONUT CHART

- Summarizes how **data usage (MB/day)** is distributed across various **device models**.

- **Visualizes** the average data usage across five mobile device models.

- Each segment size proportional to the average data usage of users of that device.



Each slice will be **labeled** with the name of the corresponding device model.

Percentages indicate the proportion of total data usage designated to each device model.

CONCLUSION

- This dataset provides valuable insights into how people interact with their devices.
- We can analyze user habits, such as app usage patterns and browsing behavior.
- The data also reveals details about device efficiency, including battery life and processing speed.
- By analyzing this data, we can:
 - Enhance mobile technology to better meet user needs.
 - Improve user experiences by optimizing device performance and creating more intuitive interfaces.

THANK YOU