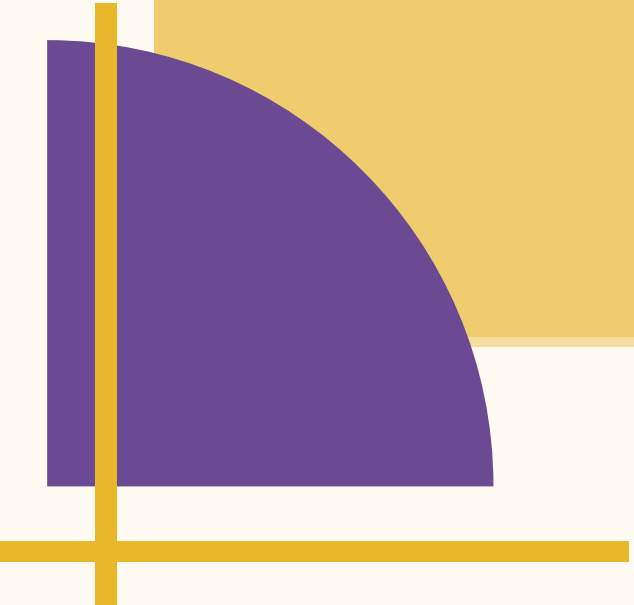
ANALYSIS OF FACEBOOK DATA



Ankita Bhanushali

Pursuing- M.Sc. Statistics

Data Analyst Consultant- Kantar

INSAID, March 2020 GCD Cohort.

Introduction About Database

Facebook, American company offering online social networking services. Facebook was founded in 2004 by Mark Zuckerberg, Eduardo Savarin, Dustin Moskovitz, and Chris Hughes, all of whom were students at Harvard University. Facebook became the largest social network in the world, with more than one billion users as of 2012, and about half that number were using Facebook every day. The company's headquarters are in Menlo Park, California. The way that Facebook works is really simple and easy to understand for users. Every user that has Facebook account can connect voluntarily with people he chooses and share content with them. This process includes sending a **"Friend request"** by the user and accepting the invitation by the recipients. The double-side feature gives every user the opportunity to decide if he wants to be **connected (or not connected)** with others. Facebook plays many vital roles in order to connect people in the world without boundary because almost every people these days are familiar with Facebook .This gives us the opportunity to know more about their culture, values, custom and tradition. Facebook's mission is to give people the power to **share** and make the world more open. People use Facebook to stay connected with friends and family, to discover what's going on in the world, and to share and express what matters to them. User can share their feelings and what's happening around in our daily life through Facebook. Facebook plays many vital roles in order to connect people in the **world** without boundary because almost every person nowadays is familiar with **Facebook** .This gives us the opportunity to know more about their culture, values, custom and tradition. They also can get feedback from their friends about their **reaction** toward user's feeling.

TABLE OF CONTENTS

01

Problem Statement

Problem Statement

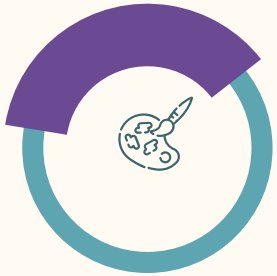
Facebook is all about giving people the power to **share** and make the world more open. People use Facebook to stay connected with **friends** and **family**, to discover what's going on in the world. User can share their **feelings** and what's happening around in our daily life through Facebook. They also can get feedback from their friends about their **reactions** toward the user's feelings.
Here we **Analyze** how people using Facebook through comparing with **Age** & **Gender**.

02

Data Profiling

- Understanding the data sets
- Pre-Profiling
- Preprocessing

Pre-Profiling



99003

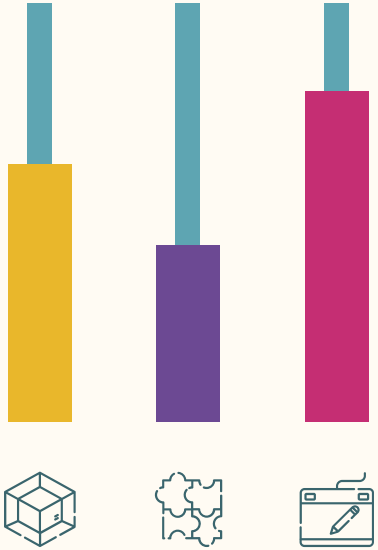
Total sample

<0.1%

Missing value

15

No. of variables



Preprocessing

Dealing with missing values

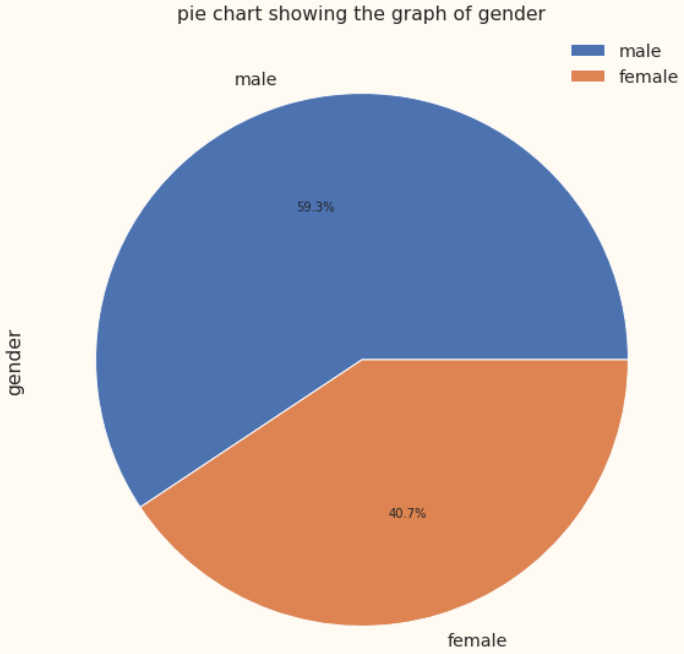
- Dropping the column **DOB Day, DOB Month, DOB Year** as they were not useful since Age column is present and considered for analysis.
- Replacing missing values of **Tenure** with median values. Since its only 0.1%.
- Replacing missing values of gender with mode values since its only 0.2%

03

Exploratory Data Analysis

- Pie chart
- Correlation Matrix
- Histogram
- Scatter plot
- Kernel Density Estimate Plot
- Bivariate Linear Chart

Pie Chart

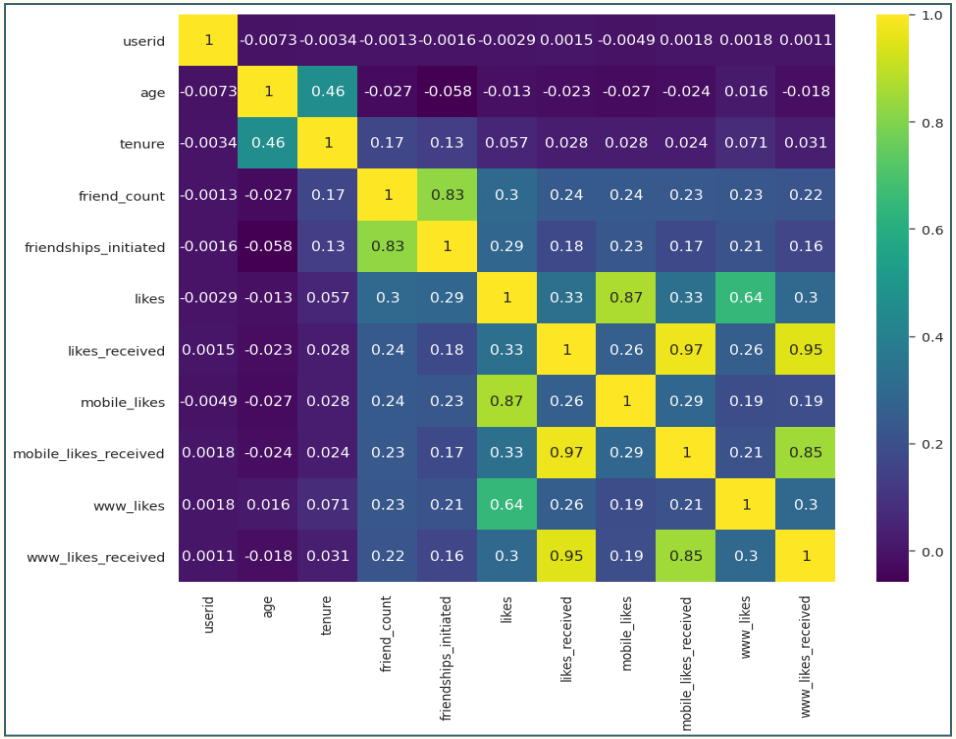


Gender:

- Male: 59.3%
- Female: 40.7%

Correlation Matrix

From correlation matrix, we observe that there exists high correlation between variables.



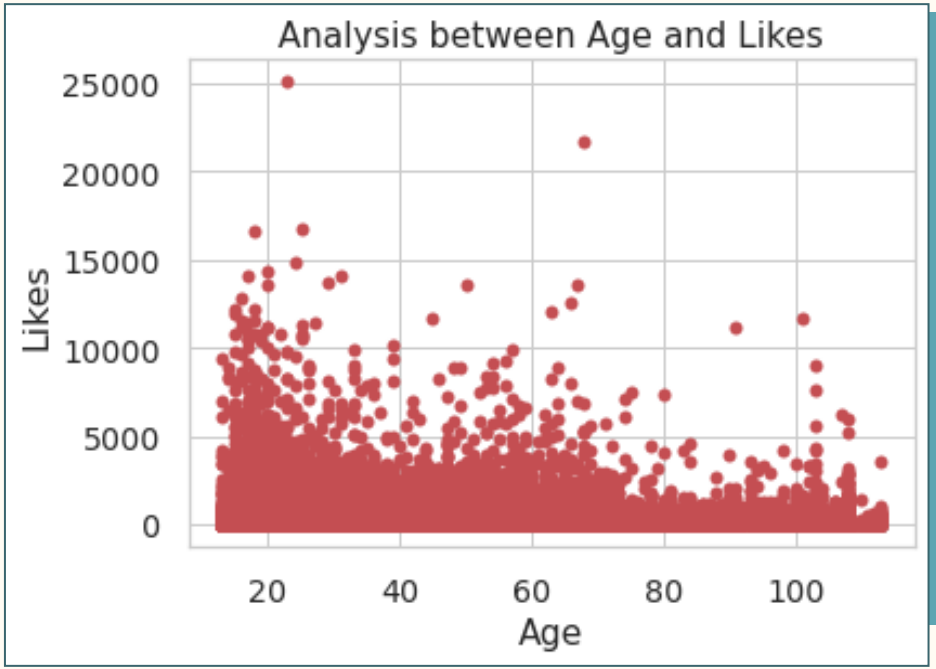
Pearson's correlation

Pearson’s correlation is -0.013

It can be seen that there is mild negative correlation between the Age and Facebook likes.

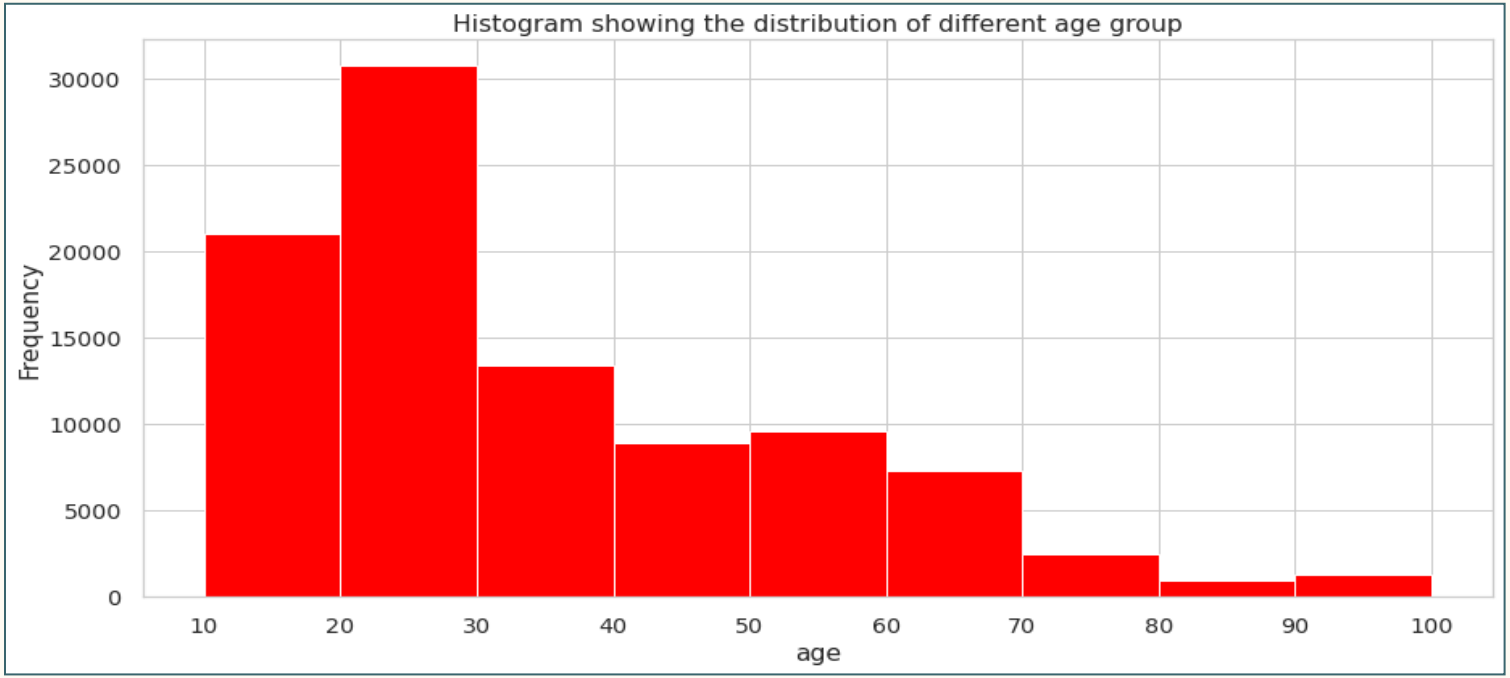
Analysis between Age and Likes

From Plot, we can see that Youngsters (Age 10-35) have more likes as compared to elders.

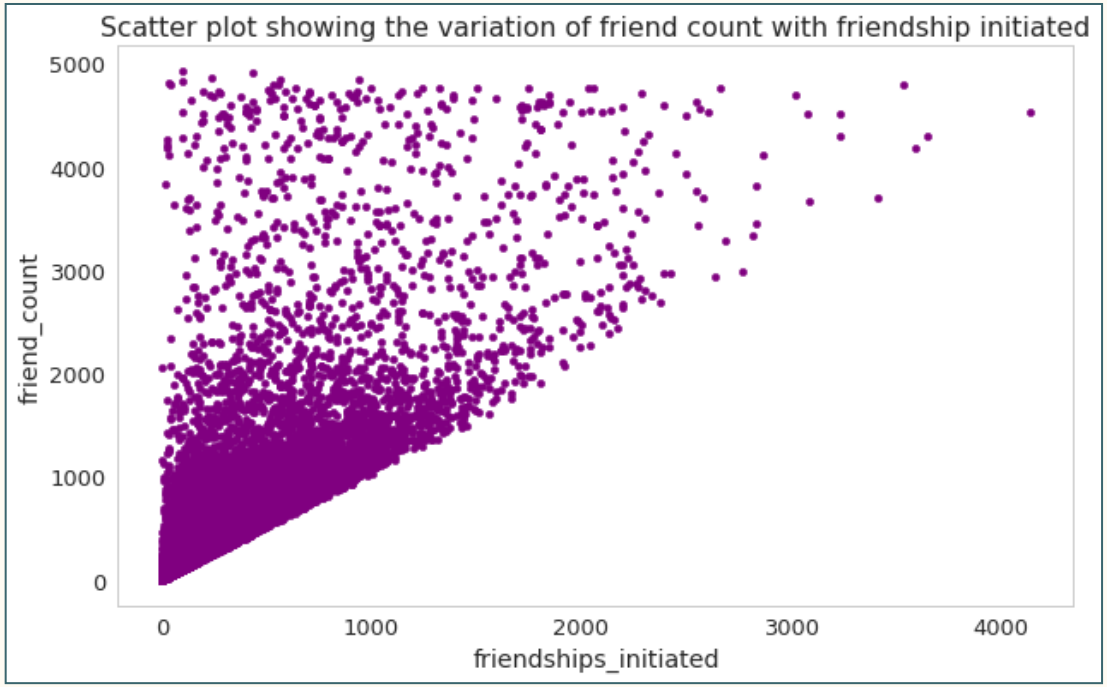


Histogram

From Histogram plot, we can see we can say that craze of Facebook is more in youngsters as compared to elders.



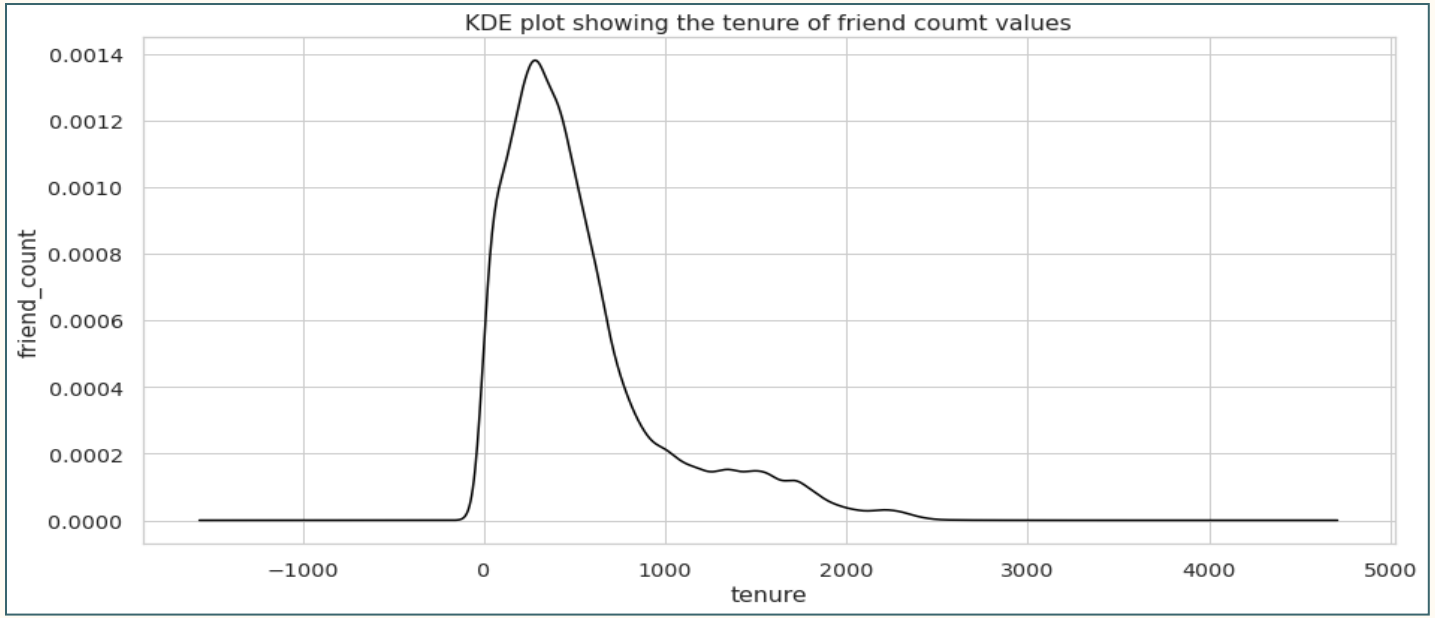
Scatter Plot



Here, We can see strong positive correlation between friends count & friendship initiated, i.e. If you initiate the friendship your friendship count increases.

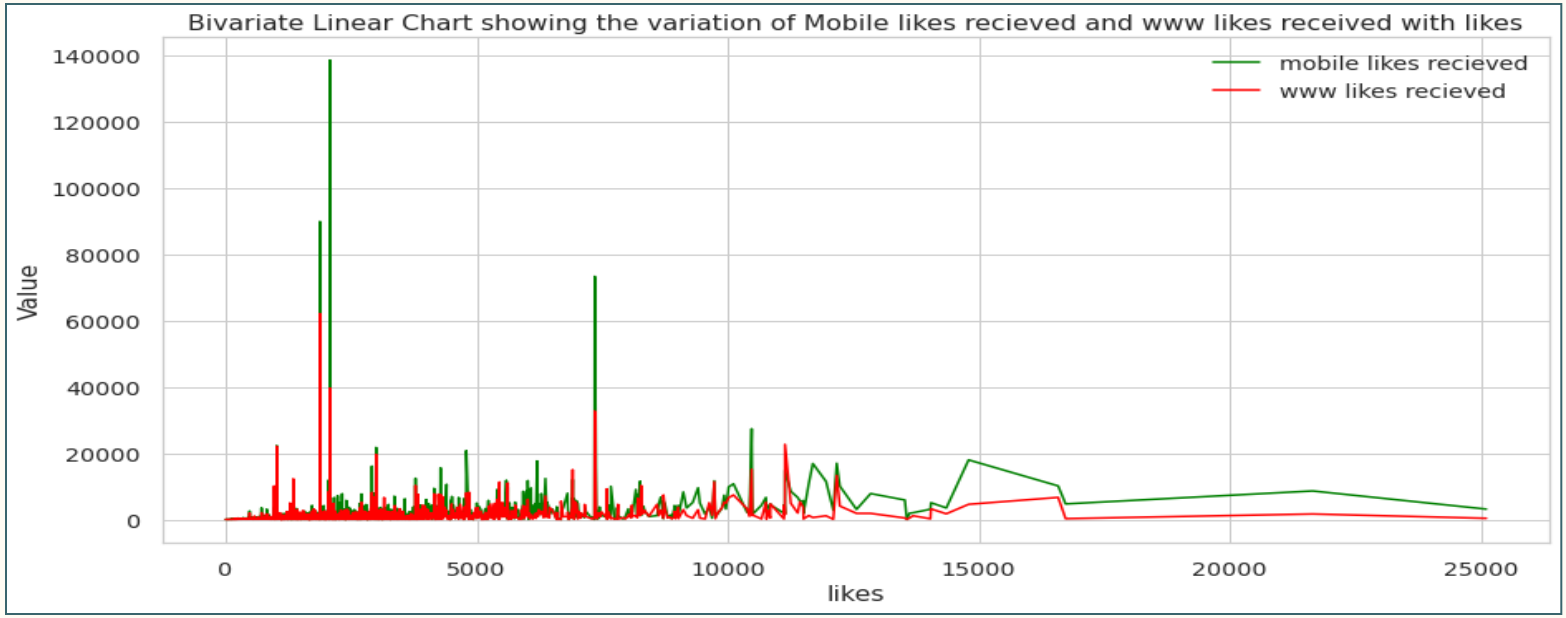
Kernel Density Estimate Plot

This density plot visualizes the distribution of data over a time period, here we can see that friendship count is more in initial days and gradually comes down after some period.



Bivariate Linear Chart

From plot, We can say that People prefer Mobile more than Website.



04

Conclusions

Conclusions

From the Analysis we can say that, **Youngsters** more attracted to social site like Facebook. They get more likes. Also their **friendship count** is more. But not only **kids** all age groups are moreover equally active on **Facebook**. friendship count is more in initial days and gradually comes down after some period. And from analysis we can also say that people prefer **mobile** more than website.

Resources

- https://www.geeksforgeeks.org/graph-plotting-in-python-set-1/
- http://www.cogsci.ucsd.edu/.
- https://app.finxter.com/learn/computer/science/
- https://stackoverflow.com

Alternative Resources

THANK YOU