

Facebook Data Science Project

Exploratory Data Analysis



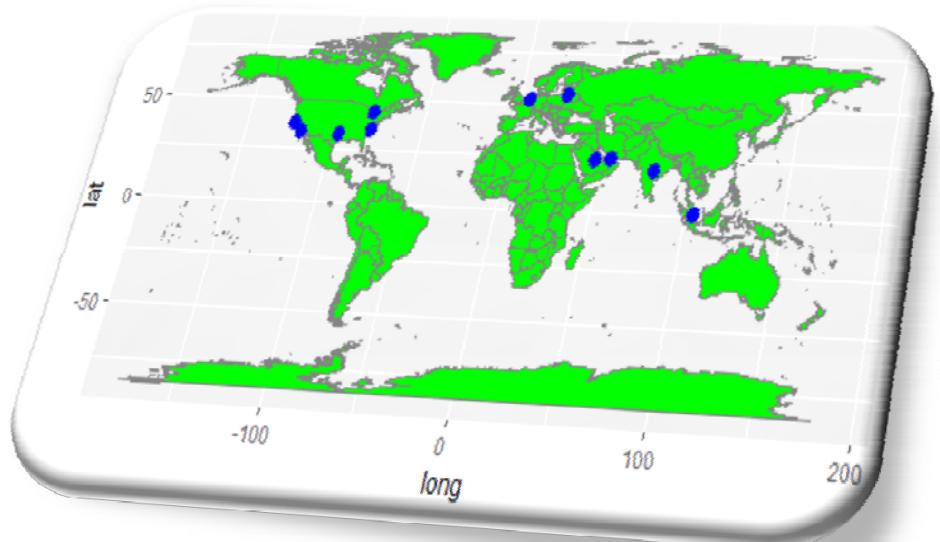
Data Diggers International Group
Feb 28, 2017


Data Diggers
Let's Uncomplicate Business



Team Info

- Ali Qaiser Syed
- Annada Prasad
- Chia Aik Lee
- Diego Jaramillo
- Kiran Devathi
- Manoj Kalamkar
- Praveen Parvataneni
- Subrata Roy
- Ravi Shanker Rao Bodkai
- Vasudev Pendyala





Domain

Social Network

Topic

Facebook Data Analysis



Introduction

Project

- This project is to analyze and explore the Facebook Data using the EDA concepts with the application of Python programming language.

Scope

- Dataset containing Facebook data

Objective

- Prepare and present the project report with HYFY Google chrome extension.





Dataset Description

- Facebook data
- Semi-structured, Multivariate data
- Time-ordered data
- Data Quality : Contains missing values
- Number of Attributes : 15
- The dataset contains polls of Facebook's male and female users among the various age groups with Likes received and Friends request accepted over the web and mobile channels.
- Source : Provided by Upx Academy for data science foundation project evaluation

Dataset Info()

# of Records	99,003
# of Columns	15

Sl.No	Column Description	Data Type
1	dob_year_dob_month_dob_day	datetime64[ns]
2	userid	int64
3	age	int64
4	gender	object
5	tenure	float64
6	friend_count	int64
7	friendships_initiated	int64
8	likes	int64
9	likes_received	int64
10	mobile_likes	int64
11	mobile_likes_received	int64
12	www_likes	int64
13	www_likes_received	int64
14	tenure_mnth	float64
15	tenure_yrs	float64

Dataset Describe()

Columns	Count	Mean	std	Min	25%	50%	75%	Max
age	99003	37.28	22.59	13	20	28	50	113
tenure	99001	537.89	457.65	0	226	412	675	3139
friend_count	99003	196.35	387.30	0	31	82	206	4923
friendships_initiated	99003	107.45	188.79	0	17	46	117	4144
likes	99003	156.08	572.28	0	1	11	81	25111
likes_received	99003	142.69	1387.92	0	1	8	59	261197
mobile_likes	99003	106.12	445.25	0	0	4	46	25111
mobile_likes_received	99003	84.12	839.89	0	0	4	33	138561
www_likes	99003	49.96	285.56	0	0	0	7	14865
www_likes_received	99003	58.57	601.42	0	0	2	20	129953
tenure_mnth	99001	1.47	1.25	0	0.62	1.13	1.85	8.6

Business Question

Does Gender and Age of users effect their likeability in Facebook?

Approach

Pie distributions of male and female gender and likes received among the genders.

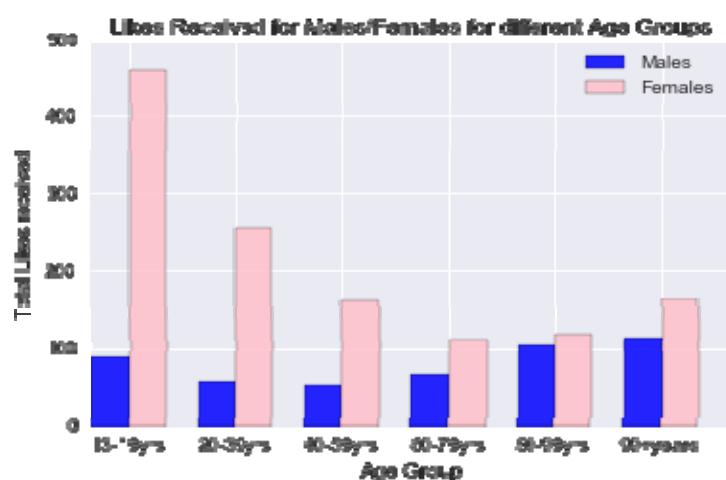
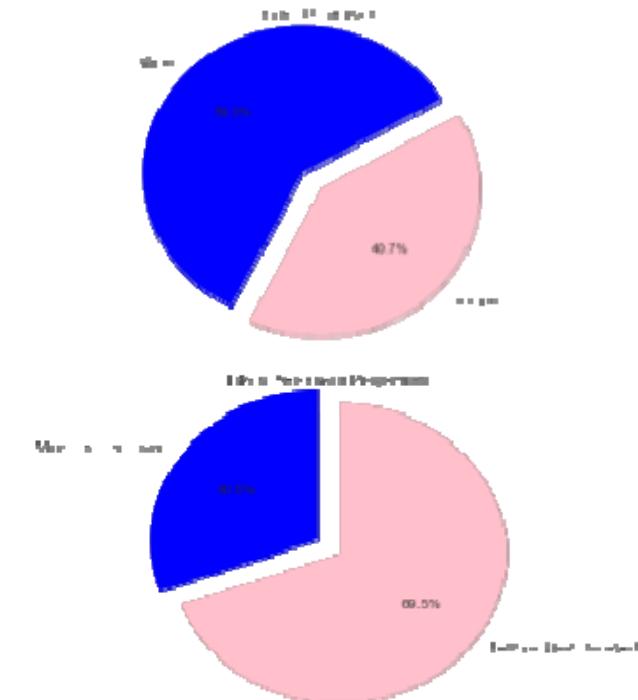
Bar Graph of Average likes received for male and females in various age groups.

Findings & Visualizations

Even though males are 19% higher in number than females, the likes received by females are more than double that of males.

Female teens got likes more than 4 times that of male teens.

In 20-39 age group also females got almost 5 times more likes than males.



Business Question

How is Facebook mostly accessed by its users and from where? **Mobile** or **Web**?

Approach

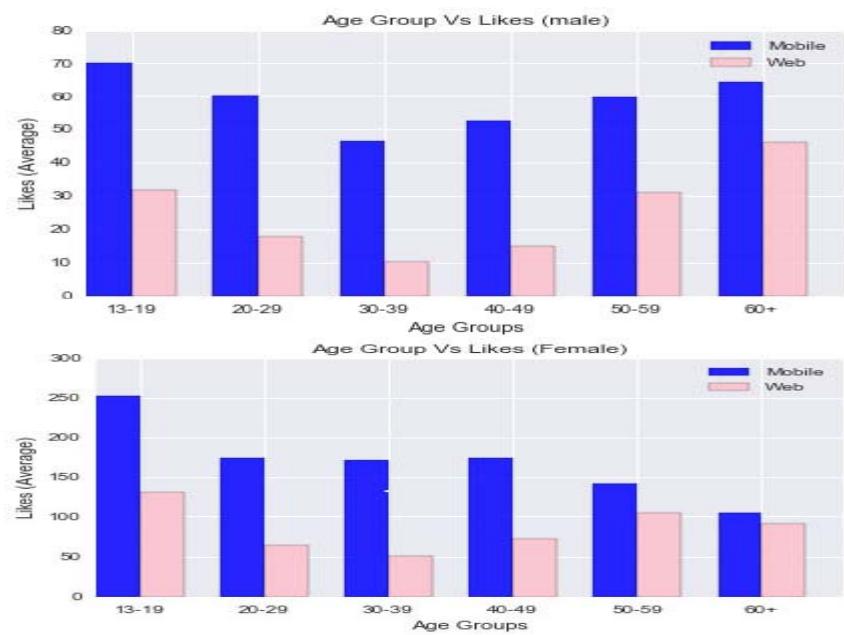
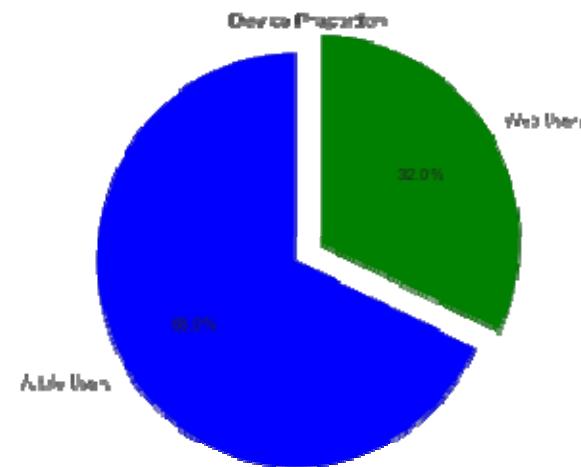
Pie distribution of Mobile and Web users accessing Facebook.

Bar Graphs of male & female users accessing Facebook on Mobile & Web channels among various age groups.

Findings & Visualizations

From the Pie distribution, it is observed that users access Facebook mostly from Mobile than Web channel.

Bar Graphs also conclude that users access Facebook largely from Mobile than Web channel across all the age groups.



Business Question

How are Likes and Likes Received related for the Facebook users?

Approach

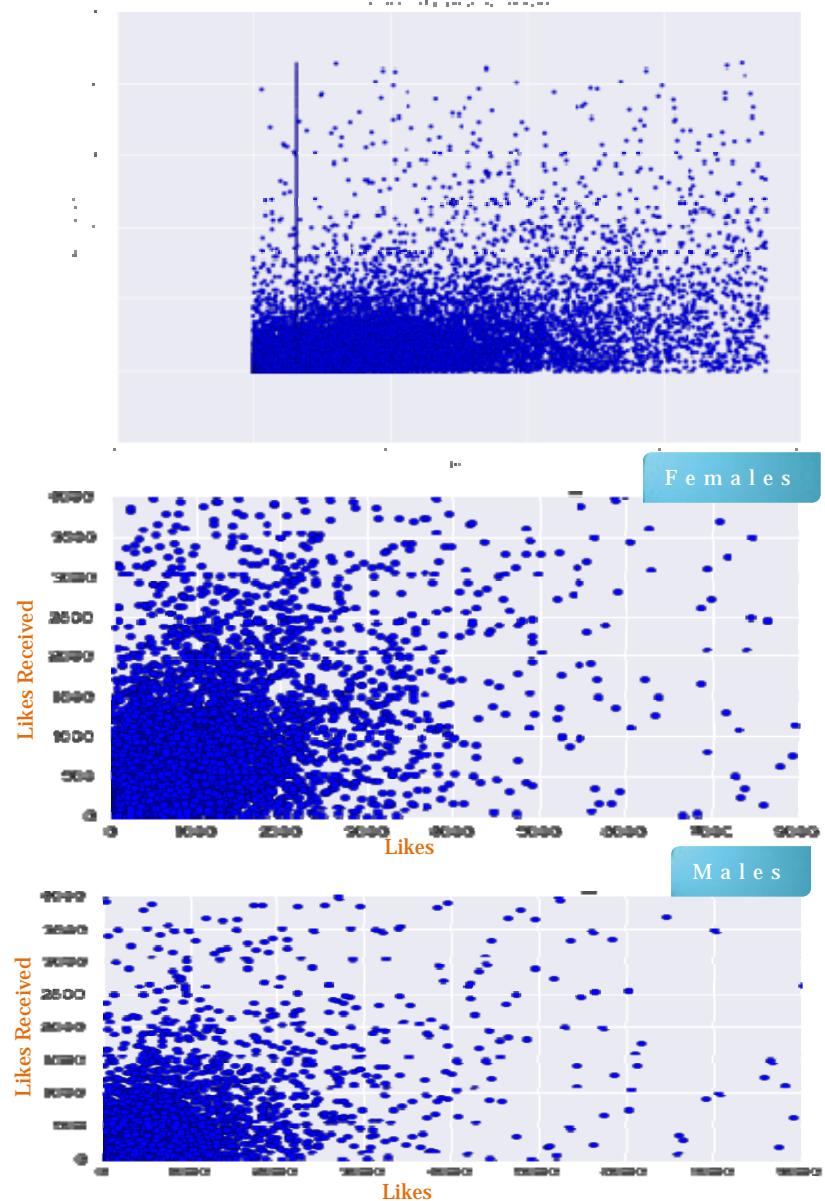
Scatter plots for the entire sample.

Scatter plots for Likes v/s Likes Received among male gender & female gender.

Findings & Visualizations

Liking More does not mean an individual be Liked Back more.

From the gender plots it is observed that females receive likes more than they give likes.



Business Question

How are Friend Count and Friendship Initiated related for the Facebook users?

Approach

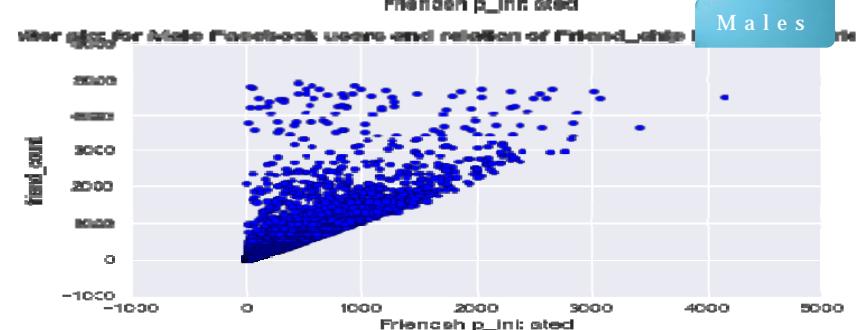
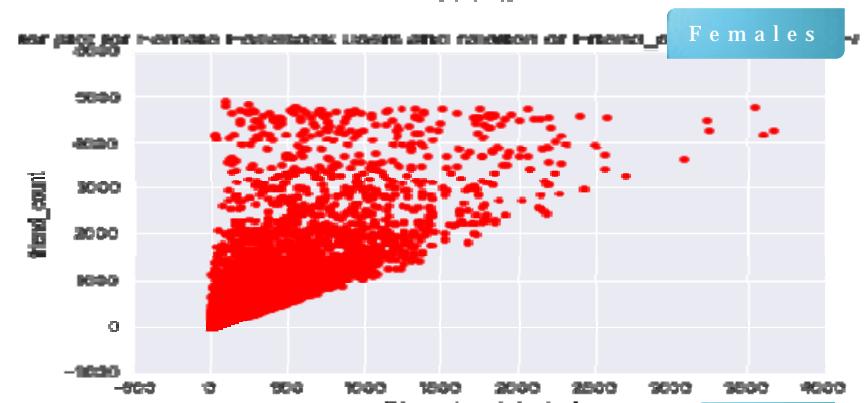
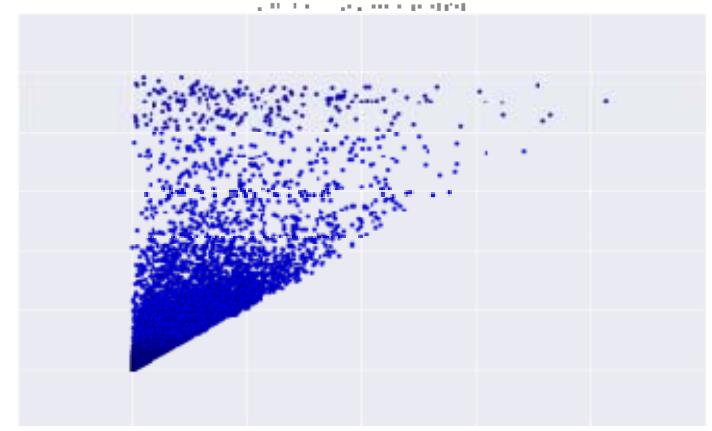
Scatter plots for the entire sample.

Scatter plots for Friend Count v/s Friendship Initiated among male gender & female gender.

Findings & Visualizations

Most users have more Friends than Friendship Initiated.

From the gender plots it is observed that females users get far more Friend requests than Friendship Initiated.



Business Question

How does **friend count** relate to users as they continue to stay long tenure on Facebook?

Approach

Scatter plot for the entire sample.

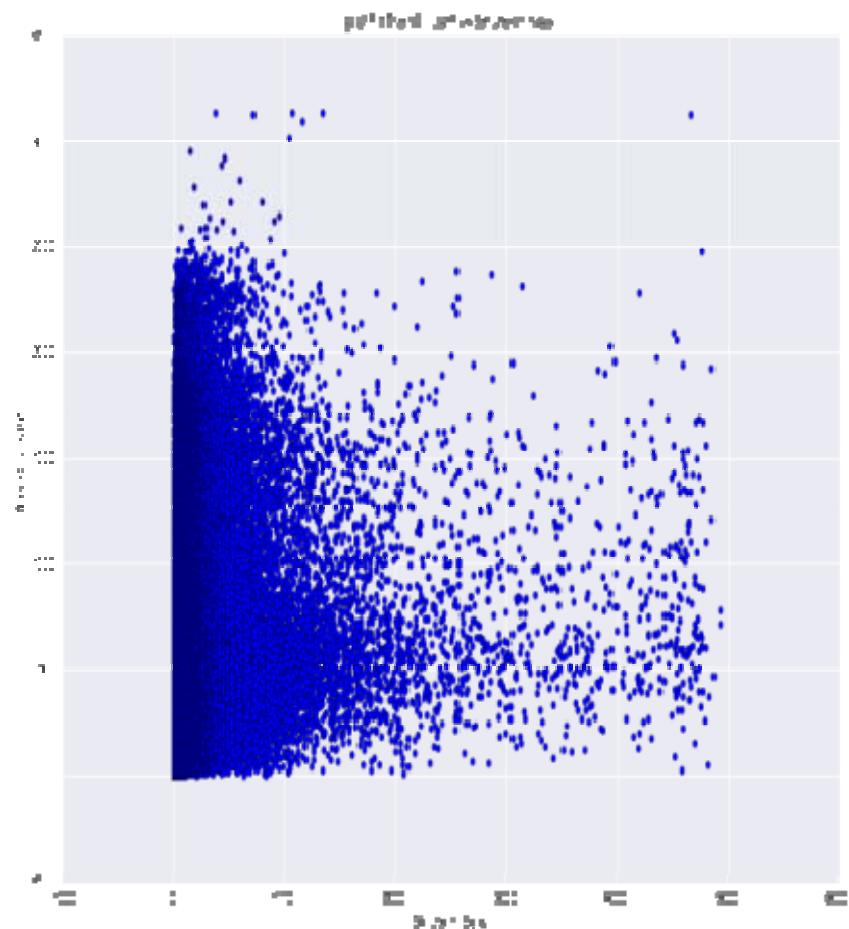
Scatter plot for Friend Count v/s Tenure among the Facebook users.

Findings & Visualizations

Longer tenure on Facebook does not translate to more friends.

Friend count rate is decreasing over time for the users.

As one gets mature it has fewer friends



Business Question

How does **likes received** relate to users as they continue to stay long tenure on Facebook?

Approach

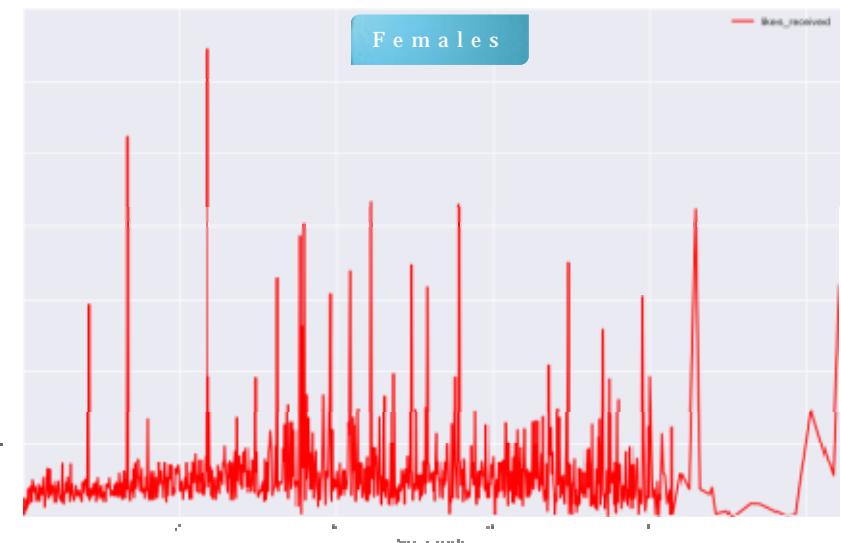
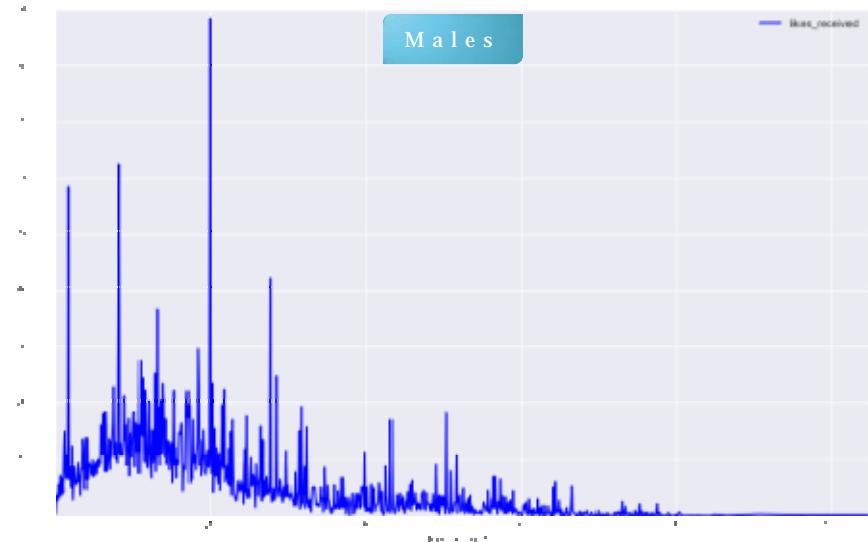
Line Graph Distribution for the entire sample.

Line Graph for Likes Received v/s Tenure (in months) among the Facebook users.

Findings & Visualizations

Likes Received for male users goes down past the 60 months tenure and keeps diminishing even more.

Likes Received for female users remained uniform up until 85 months and even beyond spikes were observed.



Business Question

How does **Zodiac Signs** impact users with their social likability?

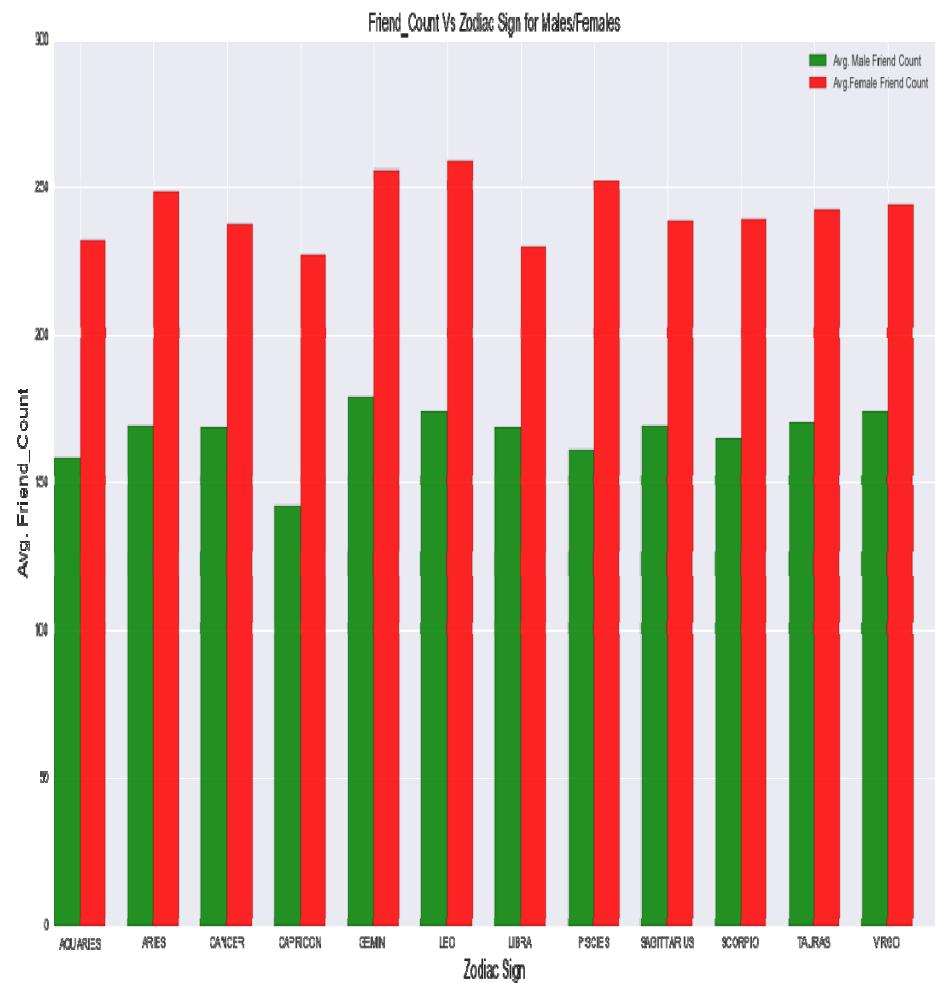
Approach

Bar Graphs for the entire sample.

Bar Graphs for Friend Count v/s Zodiac Sign among the Facebook users.

Findings & Visualizations

Friend Count (social likability) seems to be uniform among various Zodiac Signs between the male and female users on the Facebook.



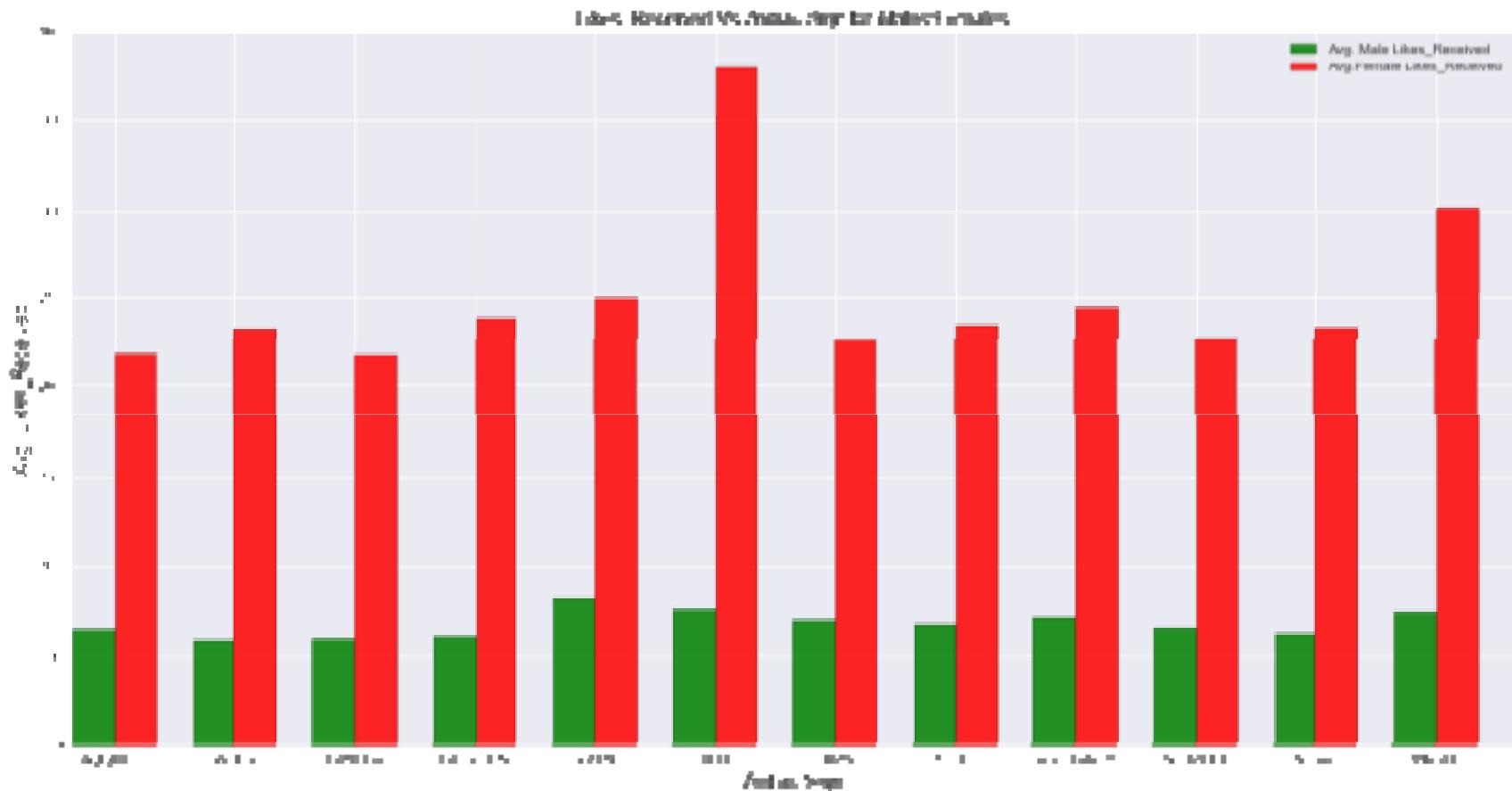
Inferential Question?

Guess why these Hollywood Stars are so popular on Social Media?

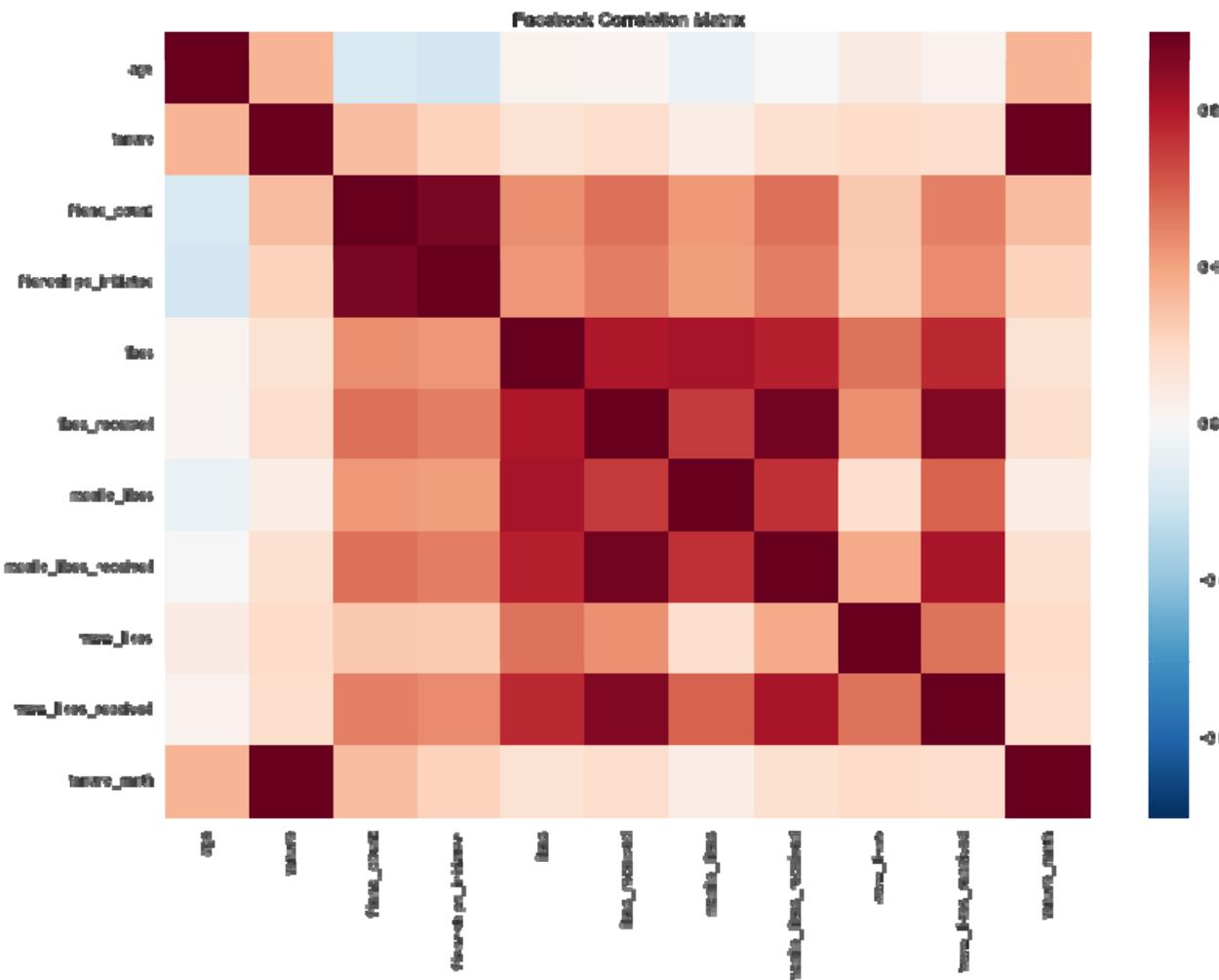


Inference

Leo Females are *more popular* followed by Virgo and others.



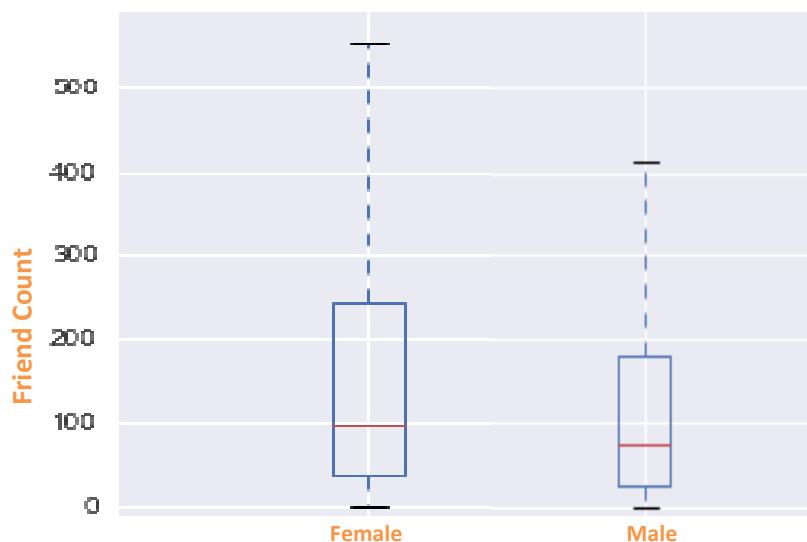
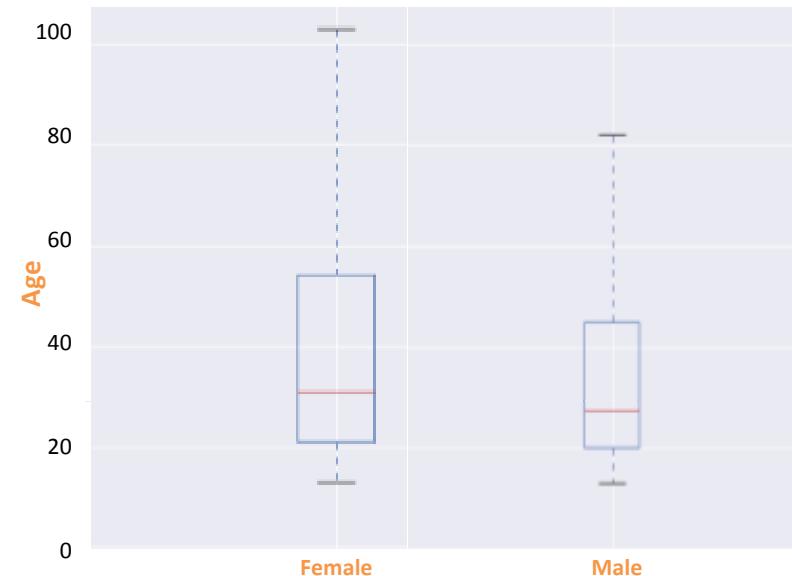
Corrogram (Correlation Heat-Map)



- positive co-relation between friend count & friendships initiated
- positive correlation for various likes received / given with total number of likes
- friend count /friendships initiated is in near to negative correlation with age

Age and Friends count by Gender

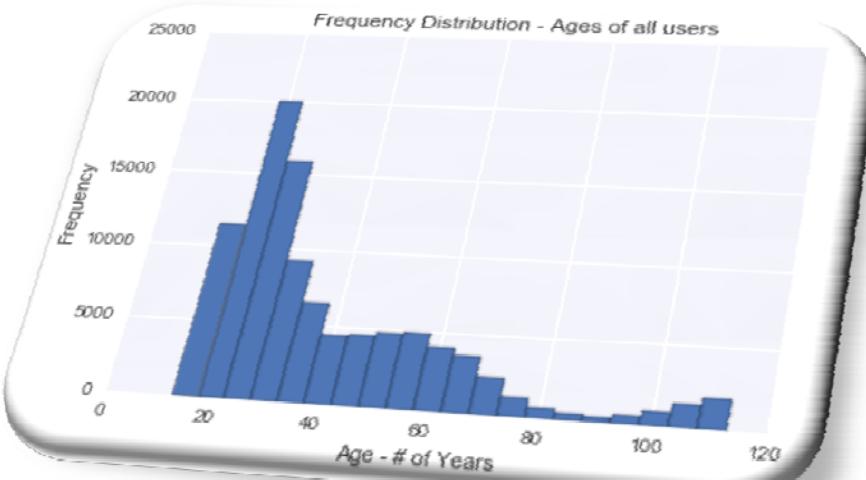
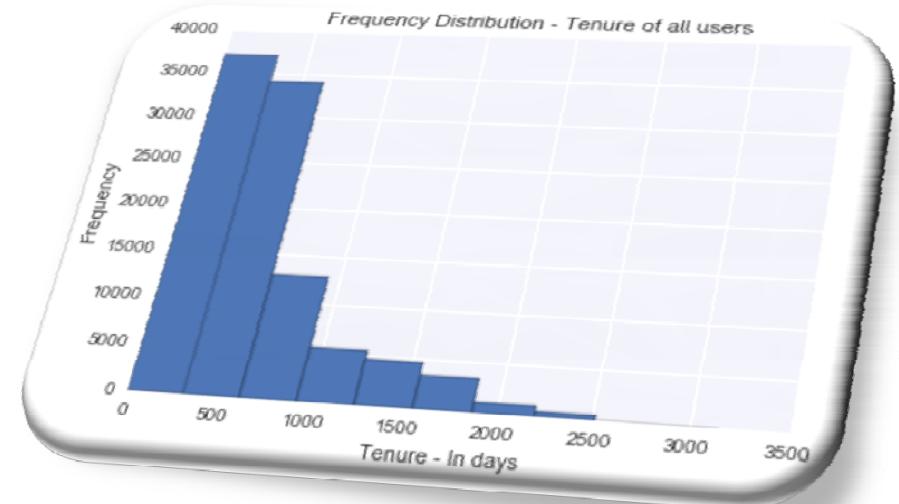
- From the **Age analysis** it is observed that average **female** age is **more** than **male** age.
- Also can observe that female users have wider age range than the male users.



- From the **Average Friend Count analysis** it is observed that **females** have **more** friends than **males**.
- Also can observe that females have wider friend count range than the males.

Facebook Tenure and Age Frequencies

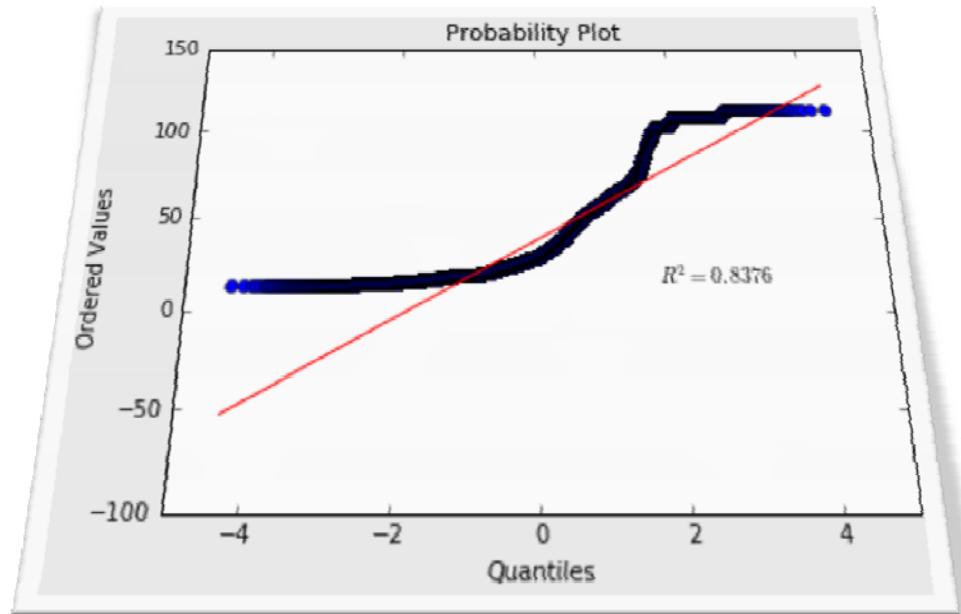
- Facebook users **Tenure frequency** analysis shows that **majority** of users joined Facebook in last couple of years (~77%).
- This could be attributed to technological transformation and Mobile First strategy where more and more users are using Mobile and hence Facebook is easily accessible.



- Facebook users **Age frequency** analysis shows that **majority** of users are of **early 20's** age and as the age grows the frequency is declining.
- Quite surprisingly frequency was increasing after **90** which raises sanctity of data for **90+** aged users. Further analysis is required to conclude that.

Normal Probability Plot

- Age pattern analysis of Facebook users
- It shows that the age is not normally distributed. However, the normal distribution here has a **high R-square** in the mid-section.



EDA – Visualization Importance

- Graph representation conveys information in more effective and efficient way than pure text description.
- Graphs representations are expressive and conveys information in a detailed way.
- Analysts can identify outliers, trend and patterns hidden in data by using appropriate visualization techniques.

Spotting outliers – ex: Box Plots

Discriminating clusters – ex: Scatter Plots

Examining relationships – ex: Correlation Plots

Comparing mean differences – ex: Cell-mean Plots

Observing a time based process – ex: Line Plots

Checking distributions – ex: Probability Plots

Thank You

*Data is the new oil
&
Analytics is the combustion engine*

