# Instagram Data Analysis Project

Eslin Gürcan 31312

## Introduction

This project analyses my art Instagram analytics (@alirex\_art). I am motivated to work on this project because I suspected number of likes a post gets might be dependent on various things, and Data Science Course Project was perfect to examine this. My aim is to select various factors and examine how these factors contribute to the likes a post has gotten.

## Data Source

Instagram allows downloading your own data as html file. My Instagram's (@alirex\_art) data is used.

## Hypothesis

Number of likes a post gets depends on the factors: month, day of the week, hour, year, tags used on the post and how many days have past since the last post before. (Explained in detail below.)

Month: Summer months get on average more likes than other months.

Day of the week: Weekends (Saturday and Sunday) gets on average more likes than other days.

Hour: Evening posts (starting from 5 pm) get less likes than other hours.

Year: 2022 on average got more likes than 2023.

Tags: Posts about Persona 5 (video game) gets more likes, therefore posts tagged with tags about Persona 5 gets on average more likes.

How many days have past since the last post increasing effects the like change from last post to the new post negatively. (Change in number of days and like change have negative correlation.)

## Data Exploration

Web-scraping was used to extract the data. Number of likes was not included in the data downloaded from Instagram. Therefore, likes were manually extracted from instagram.

Using the data graphs showing total likes and average likes depending on month, day of the week and hour of the post were plotted.

On another graphs average likes depending on month, day of the week and hour of the post throughout the years are shown.

For tags, since there were a lot of them and most of them correlated with each other for example p5, persona5, p5fanart and p5r all being about the video game persona 5, I wanted to group them. Machine learning technique NLP was used to group tags. Afterwards, average likes a group of tags got were compared to other groups.

## Data Visualisation

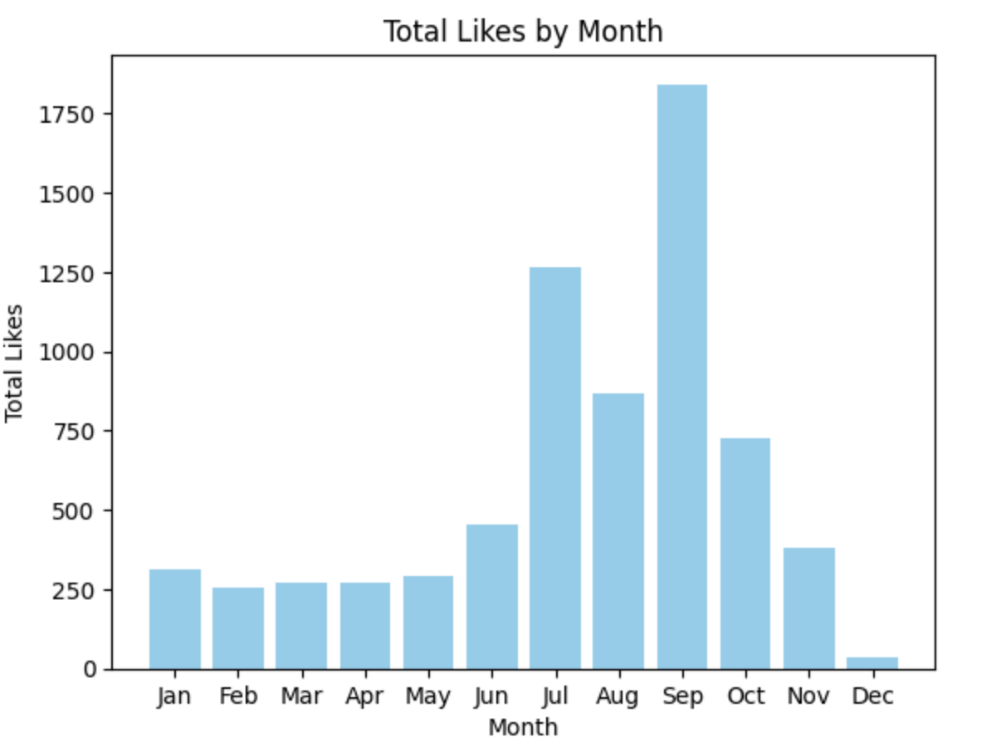


Figure 1

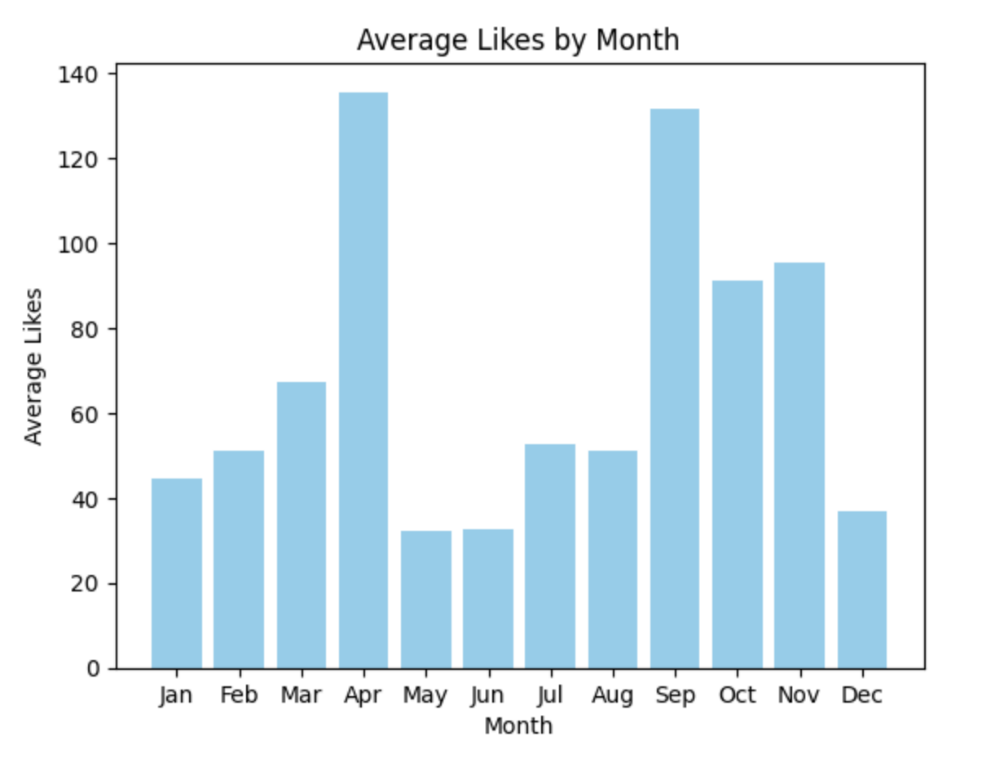
;

Figure2

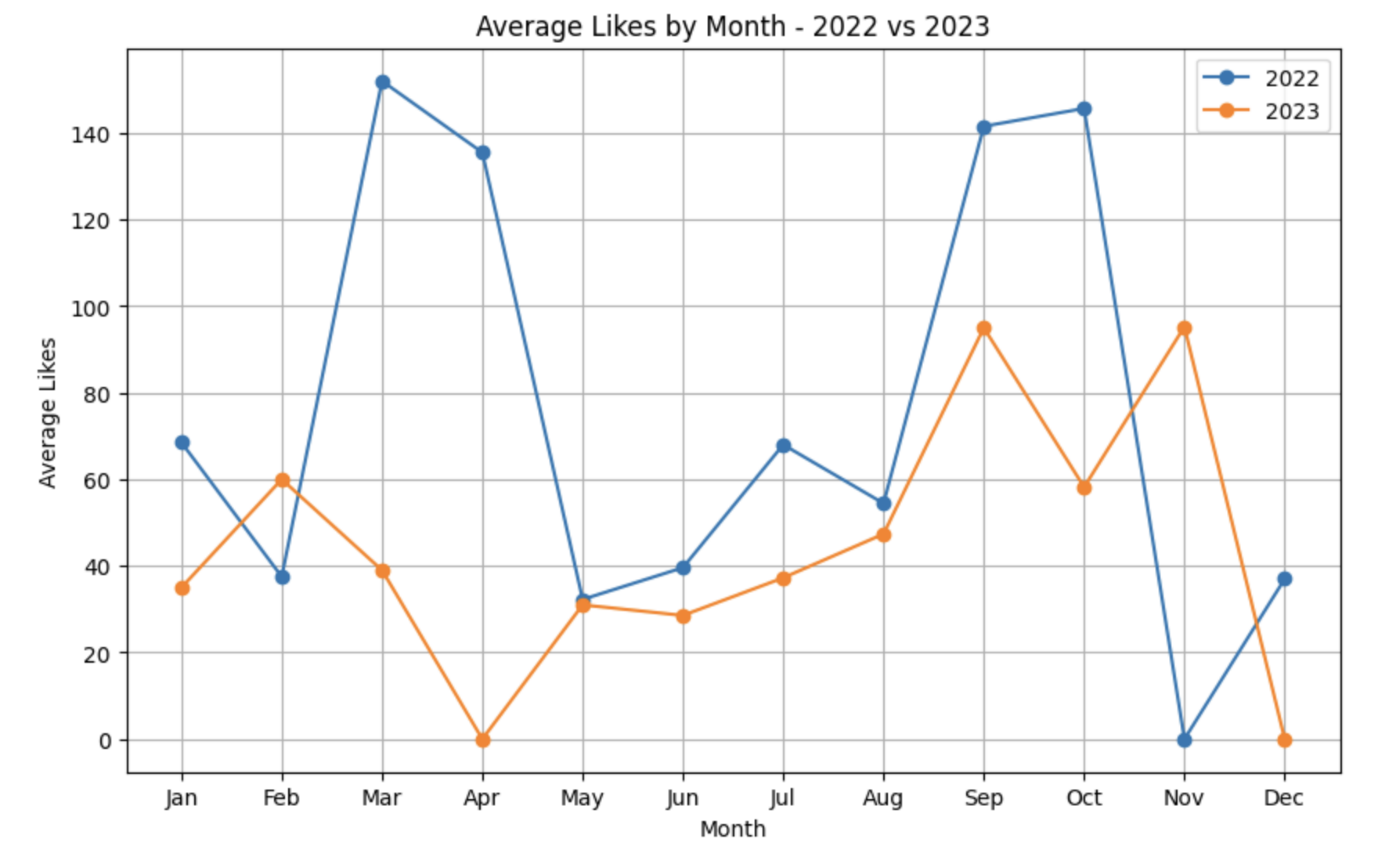


Figure 3

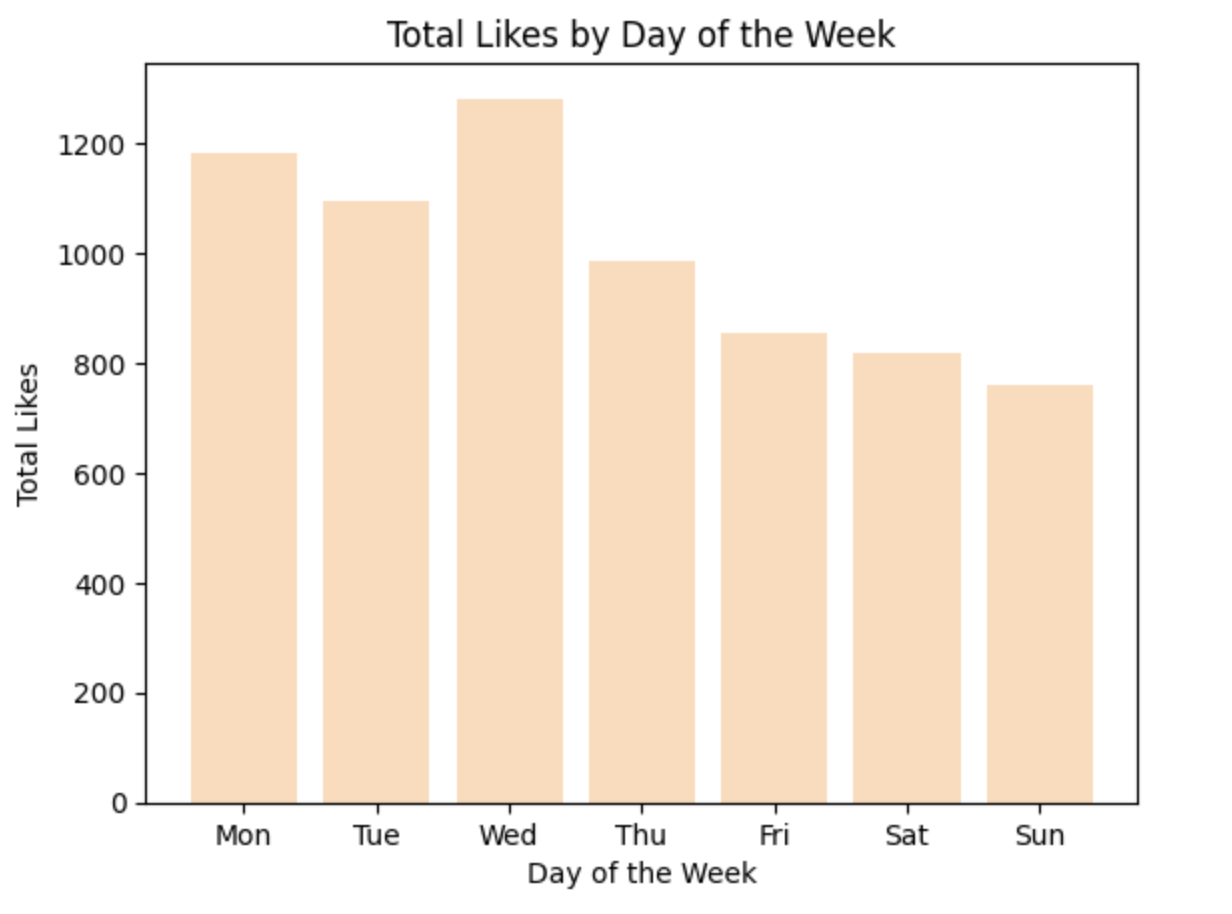


Figure 4

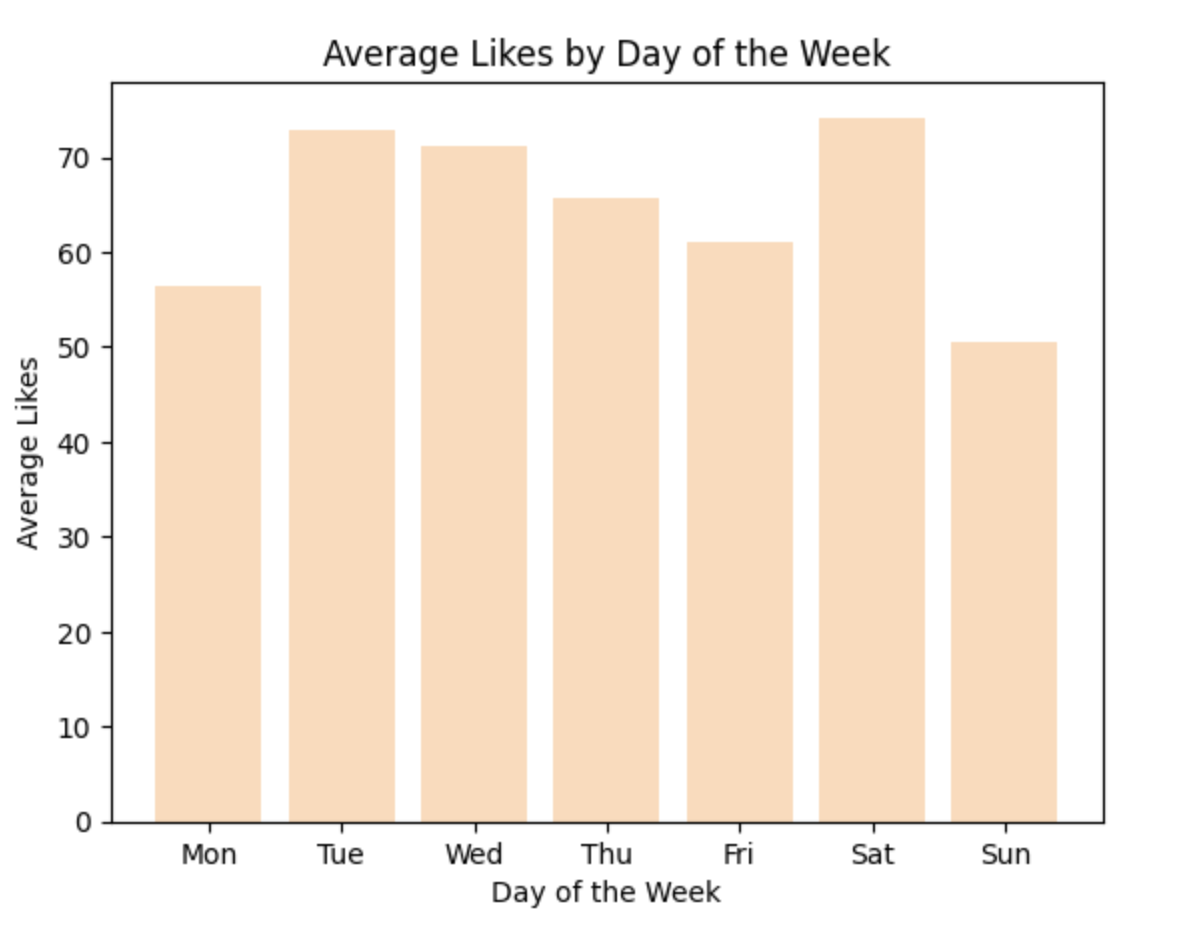


Figure 5

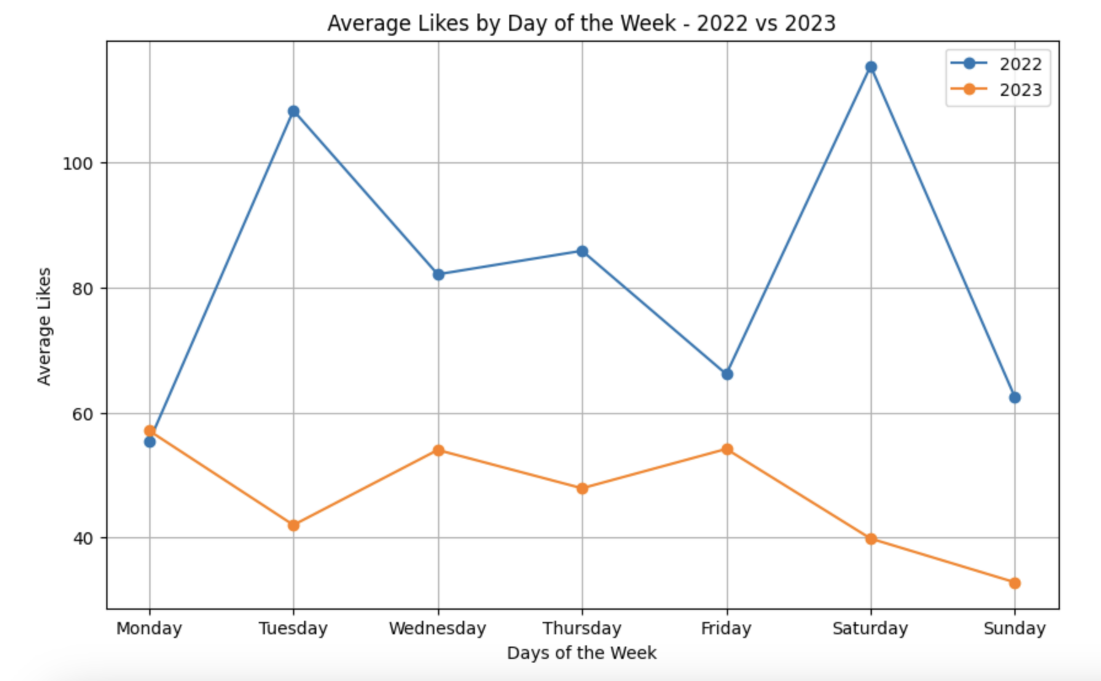


Figure 6

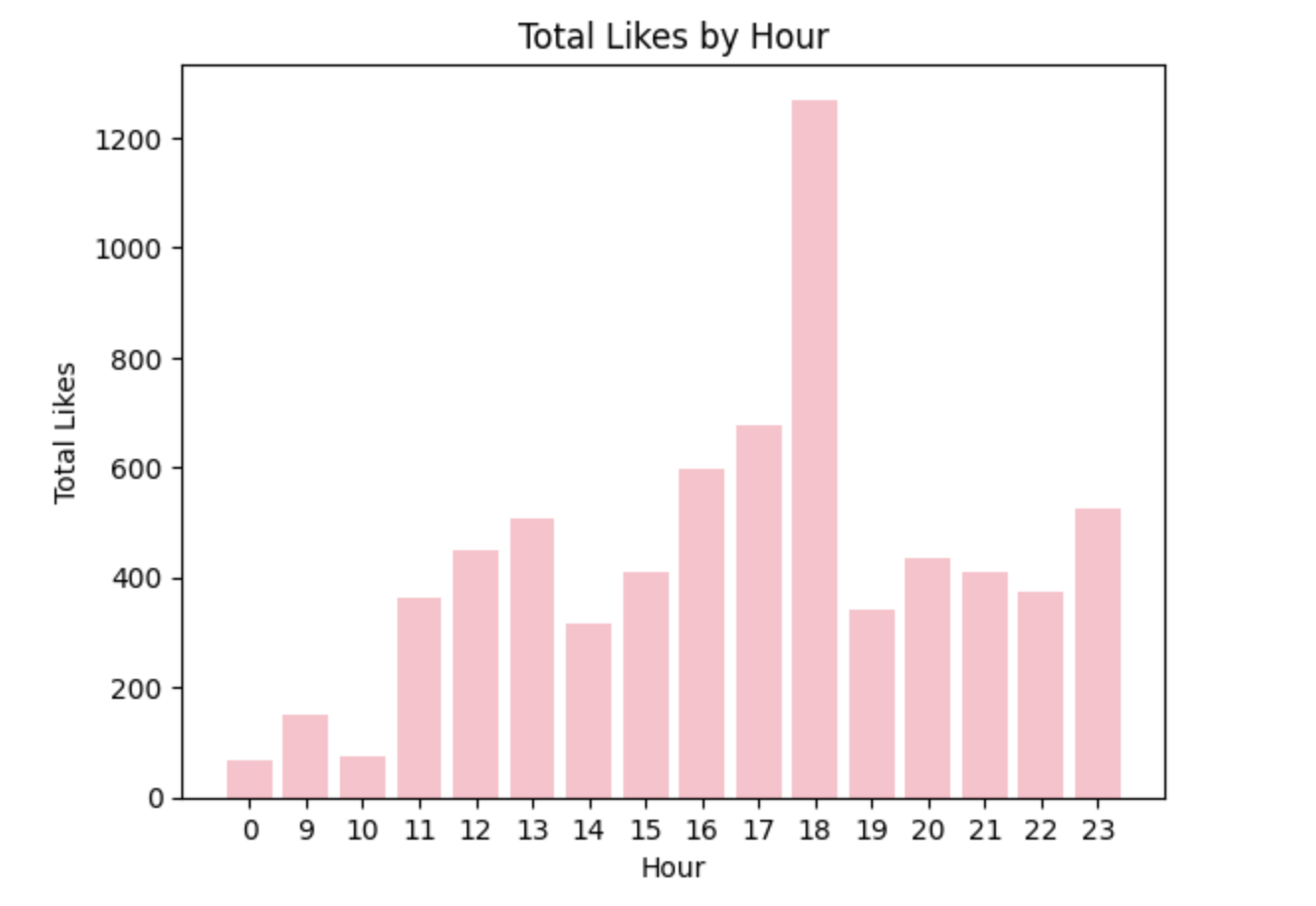


Figure 7

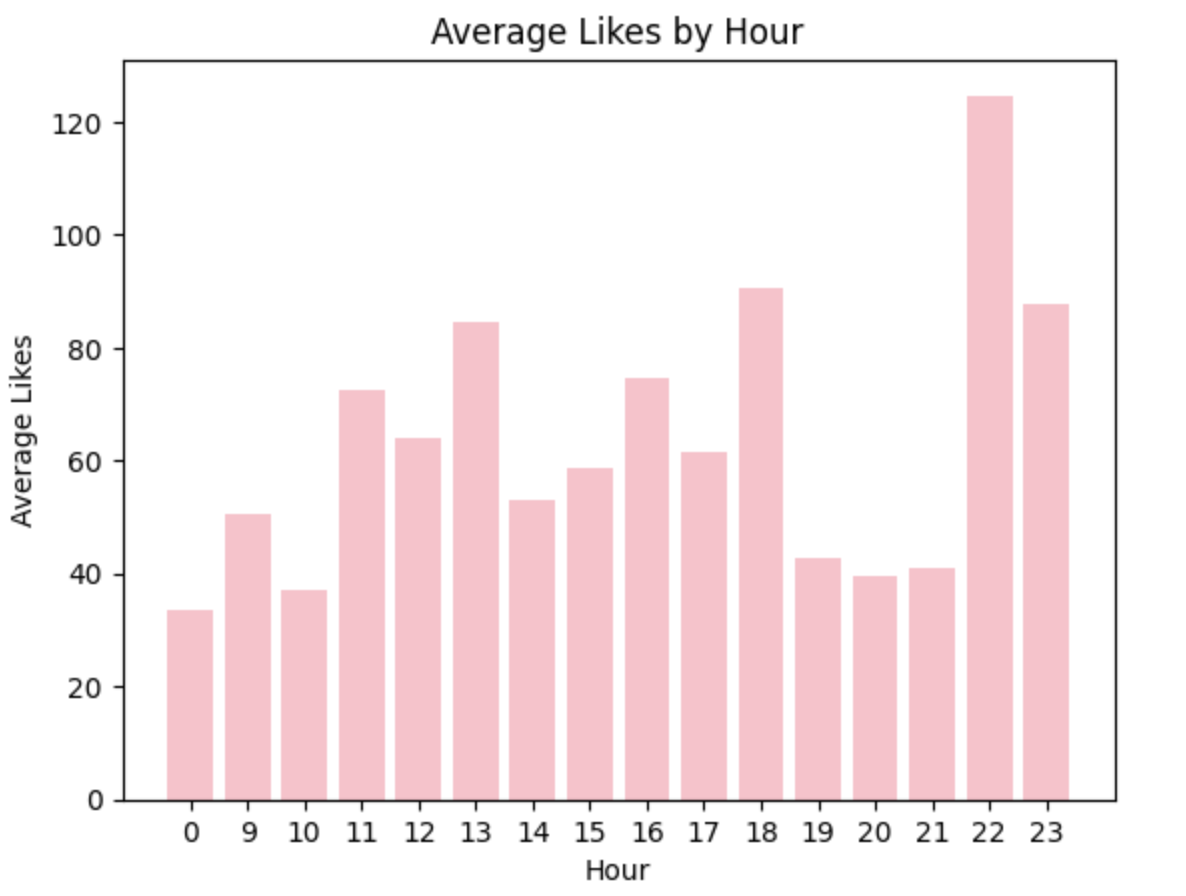


Figure 8

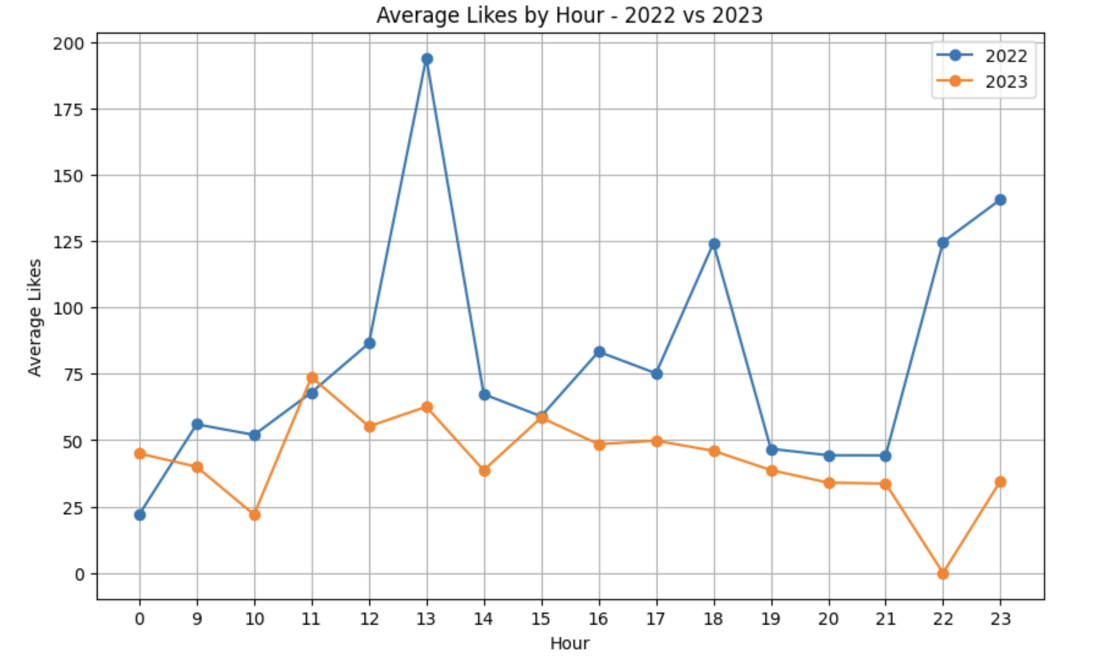


Figure 9

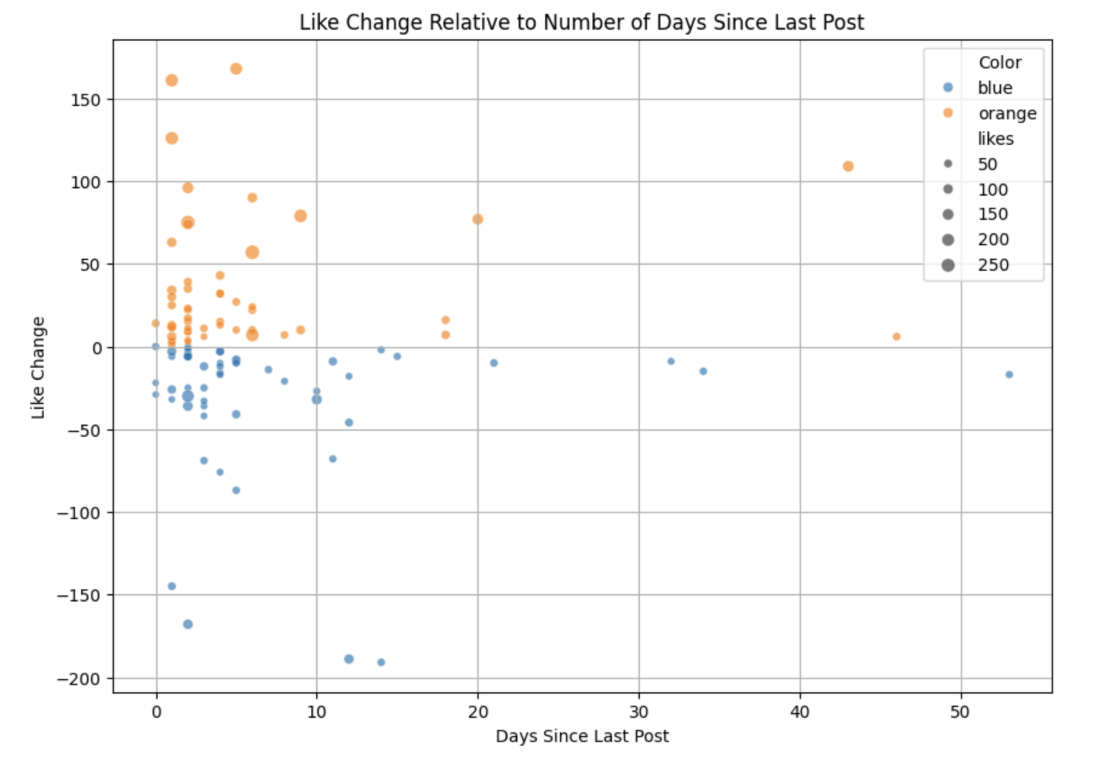


Figure 10

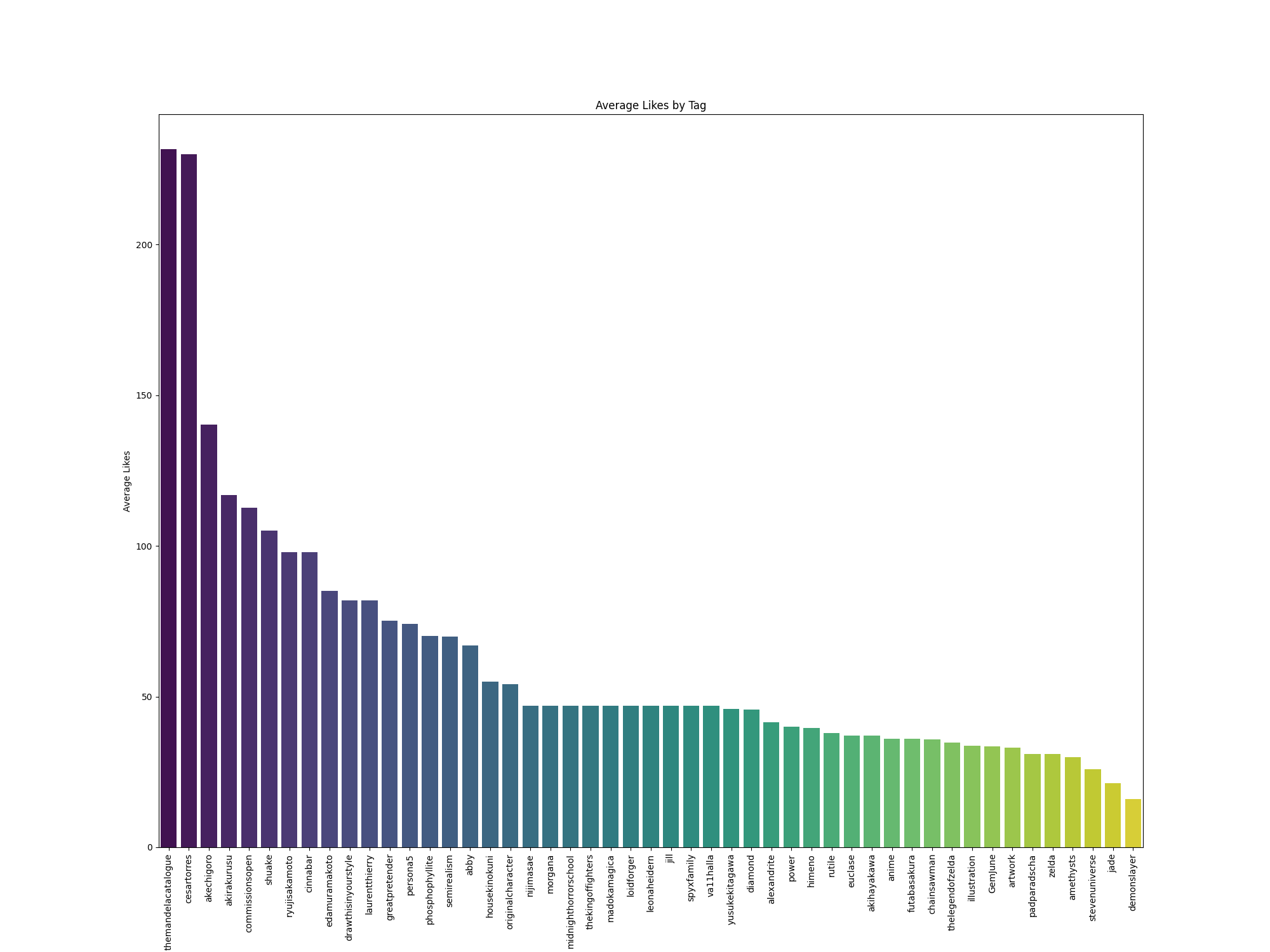


Figure 11

# Findings

Figures 1, 2 and 3 show there is a fall on average likes on May, June, July August opposed to hypothesis which suggested summer months would have higher average likes.

Figures 4, 5 and 6 indicate there is not a strong correlation with day of the week and likes as average likes for any day fluctuates between 50 and 80 likes. The graphs do not display any trend.

Figures 3 6 and 9 all show that given years 2022 and 2023; average likes for month, day of the week and hour for 2022 are higher than 2023. It can be concluded that on average 2022 likes are bigger than 2023.

Figures 8 shows average likes by hour falls on 19:00, 20:00, 21:00 (7, 8, 9 pm). This is also supported by Figure 9 where both in 2022 and 2023 average likes fall at these hours.

Figure 10 shows like difference from the last past and new post, relative to how many days have passed between them. Bigger circles, representing bigger like changes are mostly between 0 and 10 days and they are orange meaning they are bigger than 0. This implies that if a post is made after at most 10 days from the last post it is likely to gain more likes than the post before. However, there is not enough data to conclude bigger number of days between posts leads to decrease in likes.

Average likes a group of tags got was plotted on Figure 11. Mandela catalogue (analog horror series on YouTube) and a character from the show got first and second highest average likes respectively. A character from Persona 5 got the third most average likes. Persona 5 was not the tag group with highest average likes unlike the hypothesis suggested.

Conclusion: There are things that influence likes on Instagram, but findings suggest hypothesis was not entirely correct. In summary, average likes seem to fall on months May, June, July, August and at hours 7, 8, 9 pm. There was no correlation found between day of the week and likes. Year 2022 got higher average likes than 2023.

# Limitations and Future Work

Data set for this project consisted of my posts, which there were 109. I think this was not enough, for example like change relative to number of days since last post graph did not show a clear correlation.

Future Work: Mainly things concerning likes were looked at in this project. I would like to explore things about my art using image processing techniques in future. Such as which color(s) do I use the most on my art, such as which character I paint the most, can the character in the piece be predicted etc.