FOUNDATIONS OF DATA SCIENCE GROUP DIGITAL ASSIGNMENT

COURSE CODE: BCSE206L

SLOT: G1+TG1

GROUP NUMBER : **G2-08**

MEMBERS: MEHAK KHANNA 21BCE0099

ANUSHKA BHASKAR GHOSH 21BCE0504

AKSHITA KUMARI 21BCE2777

VARUN S P 21BCE3018

INSTAGRAM REACH ANALYSIS

DATA COLLECTION PHASE OF INSTAGRAM REACH ANALYSIS

INTRODUCTION TO INSTAGRAM REACH ANALYSIS

- Instagram generates a lot of data every day. When content creators post on Instagram, they do look at their engagement and reach to find the category of posts they should post more.

TYPES OF DATA TO COLLECT

Here's the Instagram Data we collected. Below are all the features in the data:

- Impressions: Number of impressions in a post (Reach)
- From Home: Reach from home
- From Hashtags: Reach from Hashtags
- From Explore: Reach from Explore
- From Other: Reach from other sources
- Saves: Number of saves
- Comments: Number of comments
- Shares: Number of shares
- Likes: Number of Likes
- Profile Visits: Numer of profile visits from the post
- Follows: Number of Follows from the post
- Caption: Caption of the post
- Hashtags: Hashtags used in the post

METHODS OF DATA COLLECTION

NATIVE PLATFORM ANALYTICS (INSIGHTS):

INSTAGRAM PROVIDES BUILT-IN ANALYTICS TOOLS KNOWN AS INSIGHTS.

ACCESSIBLE TO BUSINESS ACCOUNTS, INSIGHTS OFFER VALUABLE DATA ON POST REACH, IMPRESSIONS, PROFILE VISITS, AND AUDIENCE DEMOGRAPHICS.

FEATURES INCLUDE DATA ON FOLLOWER GROWTH, ENGAGEMENT RATE, AND TOP-PERFORMING CONTENT.

BENEFITS: CONVENIENT, INTEGRATED WITHIN THE PLATFORM, AND OFFERS COMPREHENSIVE DATA DIRECTLY FROM INSTAGRAM.

THIRD-PARTY ANALYTICS TOOLS:

VARIOUS THIRD-PARTY TOOLS LIKE SPROUT SOCIAL, HOOTSUITE, AND BUFFER OFFER ADVANCED ANALYTICS FOR INSTAGRAM.

THESE TOOLS PROVIDE ADDITIONAL FEATURES SUCH AS SCHEDULING POSTS, MONITORING MENTIONS, AND COMPETITOR ANALYSIS.

SOME TOOLS OFFER CUSTOMIZABLE REPORTS AND DEEPER INSIGHTS INTO AUDIENCE BEHAVIOR AND CONTENT PERFORMANCE.

BENEFITS: ENHANCED ANALYTICS CAPABILITIES, CROSS-PLATFORM DATA INTEGRATION, AND AUTOMATION FEATURES FOR STREAMLINED ANALYSIS.

MANUAL DATA COLLECTION:

IN SOME CASES, MANUAL DATA COLLECTION MAY BE NECESSARY, PARTICULARLY FOR QUALITATIVE DATA OR SPECIFIC METRICS NOT READILY AVAILABLE THROUGH AUTOMATED TOOLS.

THIS INVOLVES RECORDING DATA MANUALLY INTO SPREADSHEETS OR DATABASES, SUCH AS TRACKING HASHTAG USAGE, MONITORING COMPETITOR ACTIVITY, OR CONDUCTING AUDIENCE SURVEYS.

MANUAL COLLECTION REQUIRES MORE TIME AND EFFORT BUT CAN OFFER VALUABLE INSIGHTS INTO SPECIFIC ASPECTS OF INSTAGRAM REACH NOT CAPTURED BY AUTOMATED TOOLS.

BENEFITS: FLEXIBILITY TO GATHER CUSTOM DATA, DEEPER QUALITATIVE INSIGHTS, AND ABILITY TO SUPPLEMENT AUTOMATED ANALYTICS WITH ADDITIONAL INFORMATION.

DATA COLLECTED BY US

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visi	Follows	Caption	Hashtags			
2	3920	2586	1028	619	56	98	9	5	162	35	2	Here are some of the most impo	#finance #money #business #investing			
3	5394	2727	1838	1174	78	194	7	14	224	48	10	Here are some of the best data	#healthcare #health #covid #data #da			
4	4021	2085	1188	0	533	41	11	1	131	62	12	Learn how to train a machine le	#data #datascience #dataanalysis #da			
5	4528	2700	621	932	73	172	10	7	213	23	8	Here's how you can write a Pyth	#python #pythonprogramming #pytho			
6	2518	1704	255	279	37	96	5	4	123	8	0	Plotting annotations while visua	#datavisualization #datascience #data			
7	3884	2046	1214	329	43	74	7	10	144	9	2	Here are some of the most impo	#data #datascience #dataanalysis #da			
8	2621	1543	599	333	25	22	5	1	76	26	0	Learn how to analyze a candlest	#stockmarket #investing #stocks #trad			
9	3541	2071	628	500	60	135	4	9	124	12	6	Here are some of the best book	#python #pythonprogramming #pytho			
10	3749	2384	857	248	49	155	6	8	159	36	4	Here are some of the best data	#dataanalytics #datascience #data #m			
11	4115	2609	1104	178	46	122	6	3	191	31	6	Here are two best ways to coun	#python #pythonprogramming #pytho			
12	2218	1597	411	162	15	28	6	3	81	29	4	Learn the implementation of Ale	#neuralnetwork #machinelearning #ar			
13	3234	2414	476	185	75	122	8	14	151	15	0	Here's how to get the live stock	#python #pythonprogramming #pytho			
14	4344	2168	1274	673	40	119	7	11	162	8	2	Here are some of the most impo	#data #datascience #dataanalysis #da			
15	3216	2524	212	201	223	121	5	5	142	20	4	Here's how to visualize an intera	#stockmarket #investing #stocks #trad			
16	9453	2525	5799	208	794	100	6	10	294	181	42	Omicron Variant Sentiment Ana	#data #datascience #dataanalysis #da			
17	5055	2017	2351	298	108	101	7	11	159	17	6	In Data Science, Time Series Ana	#timeseries #time #statistics #datascie			
18	4002	3401	278	128	73	111	17	18	205	16	2	Here are some of the highest pa	#career #job #jobs #jobsearch #educa			
19	3169	1979	707	341	32	106	8	1	121	21	2	In Data Science, Time Series Ana	#timeseries #time #statistics #datascie			
20	6168	2177	3450	153	296	82	6	6	151	77	30	Stress, anxiety, and depression a	#data #datascience #dataanalysis #da			
21	2407	1338	655	276	39	40	8	20	72	10	0	Data Science Use Cases: Here's	#data #datascience #dataanalysis #da			
22	2064	1304	362	249	37	49	4	5	76	9	0	A boxplot is a statistical data vis	#datavisualization #datascience #data			
23	3973	2415	745	676	18	72	3	4	91	11	24	Activation Functions are functio	#neuralnetwork #machinelearning #ar			
24	7281	3065	1254	2081	748	167	7	9	195	144	100	Learn how to send automatic er	#python #pythonprogramming #pytho			
25	3052	2608	201	121	87	63	5	14	129	14	2	Here are all the programming la	#programming #coding #programmer			
26	4628	2406	1260	861	26	144	8	3	160	10	4	Deep Learning is a subset of ma	#deeplearning #machinelearning #artif			
27	4082	2195	1248	541	36	135	11	5	175	19	6	Deep learning is a subset of mac	#deeplearning #machinelearning #artif			
28	5394	2275	2075	45	65	61	10	6	147	69	6	You must have heard or investe	#data #datascience #dataanalysis #da			

BEST PRACTICES FOR DATA COLLECTION

- Set specific goals and metrics to track
- Regularly monitor insights and analytics
- Utilize a combination of quantitative and qualitative data
- Maintain consistency in data collection methods
- Ensure compliance with Instagram's terms of service and privacy policies

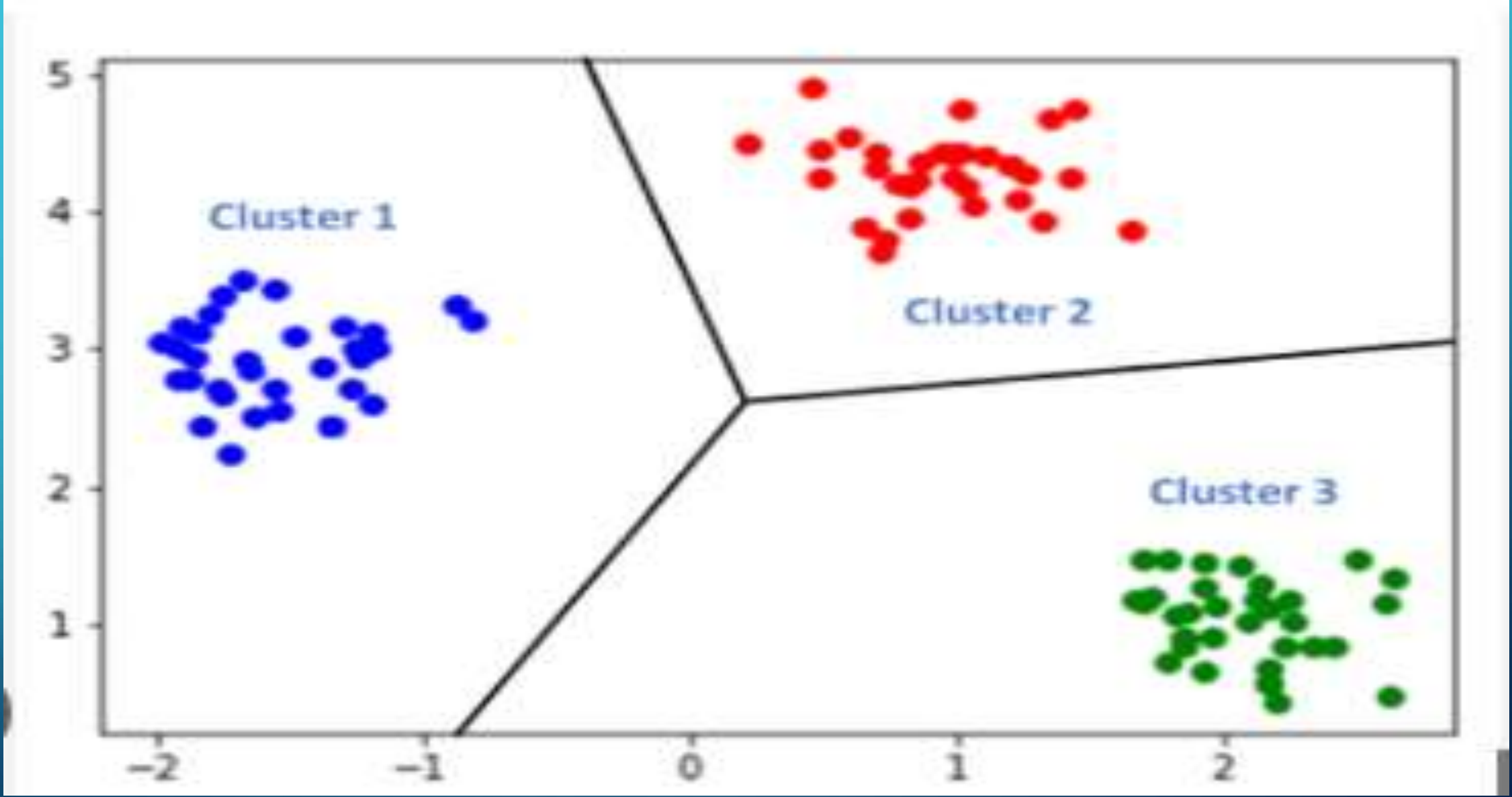
CLEANING THE DATA

CLEANING THE DATA IS AN ESSENTIAL PREPROCESSING STEP IN ANY DATA ANALYSIS OR MACHINE LEARNING PROJECT. IT INVOLVES IDENTIFYING AND RECTIFYING ERRORS, INCONSISTENCIES, AND MISSING VALUES WITHIN THE DATASET TO ENSURE ITS QUALITY AND RELIABILITY. COMMON DATA CLEANING TASKS INCLUDE REMOVING DUPLICATES, HANDLING MISSING VALUES THROUGH IMPUTATION OR DELETION, CORRECTING ERRORS IN DATA ENTRY, AND DETECTING AND ADDRESSING OUTLIERS. BY CLEANING THE DATA, ANALYSTS AND DATA SCIENTISTS CAN ENSURE THAT THEIR MODELS ARE BUILT ON ACCURATE AND REPRESENTATIVE DATA, LEADING TO MORE RELIABLE INSIGHTS AND BETTER PERFORMANCE OF THE MACHINE LEARNING ALGORITHMS. EFFECTIVE DATA CLEANING PRACTICES ARE CRUCIAL FOR PRODUCING ACTIONABLE AND TRUSTWORTHY RESULTS IN DATA-DRIVEN DECISION-MAKING PROCESSES.


Clustering
Ignore the tuples
Methods of Data Cleaning
Regression
Fill the missing value
Binning

CLUSTERING

- Clustering in the context of data cleaning refers to the process of grouping similar data points together based on certain characteristics or features. This technique is often used to identify and handle outliers, detect anomalies, and partition the data into more manageable subsets for further analysis or processing. In data cleaning, clustering algorithms can be applied to identify groups of data points that exhibit similar patterns or behaviors, allowing analysts to identify potential errors, inconsistencies, or outliers within the dataset. By clustering the data, analysts can gain insights into the underlying structure of the dataset and make informed decisions about how to handle and clean the data effectively. Clustering techniques play a valuable role in data cleaning by providing a systematic approach to identifying and addressing data quality issues in large and complex datasets.


Cluster 1
Cluster 2
Cluster 3
5
4
3
2
1
−2
−1
0
1
2

IGNORING THE TUPLES

- Ignoring tuples in data cleaning refers to the process of excluding or disregarding certain rows or observations (tuples) from the dataset during the cleaning process. This could be done for various reasons, such as if the tuples contain missing or incomplete information, outliers, or if they are deemed irrelevant to the analysis. Ignoring tuples can help improve the quality and reliability of the dataset by removing instances that could potentially introduce bias or inaccuracies into the analysis. However, it's essential to carefully consider the implications of ignoring tuples and ensure that it does not adversely affect the integrity or representativeness of the dataset for the intended analysis or modeling tasks.

id	first_name	last_name	email
1	Carine	Schmitt	carine.schmitt@verizon.net
4	Janine	Labrune	janine.labrune@aol.com
6	Janine	Labrune	janine.labrune@aol.com
2	Jean	King	jean.king@me.com
12	Jean	King	jean.king@me.com
5	Jonas	Bergulfsen	jonas.bergulfsen@mac.com
10	Julie	Murphy	julie.murphy@yahoo.com
11	Kwai	Lee	kwai.lee@google.com
3	Peter	Ferguson	peter.ferguson@google.com
9	Roland	Keitel	roland.keitel@yahoo.com
14	Roland	Keitel	roland.keitel@yahoo.com
7	Susan	Nelson	susan.nelson@comcast.net
13	Susan	Nelson	susan.nelson@comcast.net
8	Zbyszek	Piestrzeniewicz	zbyszek.piestrzeniewicz@att.net

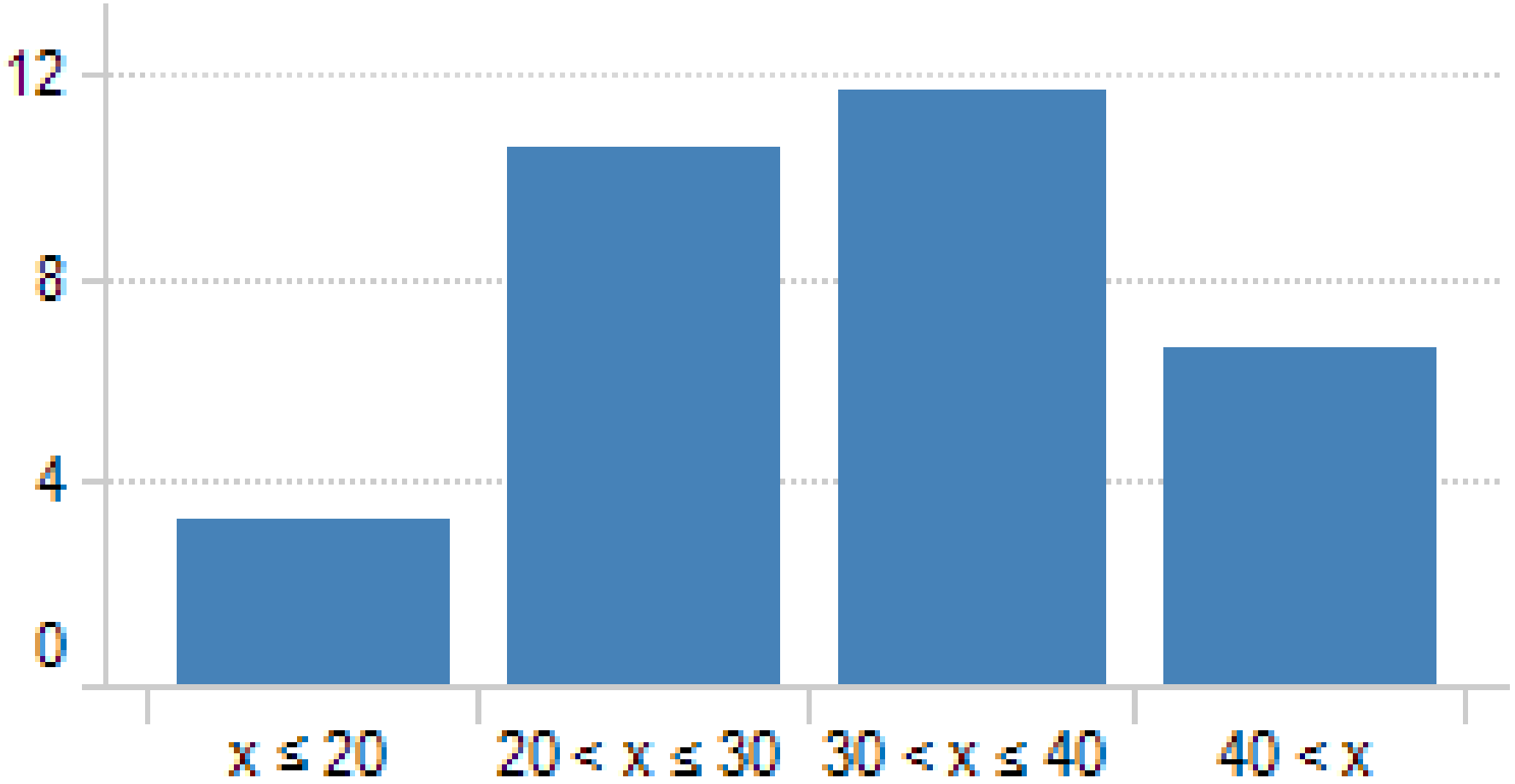
FILLING THE MISSING VALUE

- In ChatGPT, filling missing values typically involves predicting or generating plausible values to replace the missing ones in a conversation or text. ChatGPT can accomplish this by leveraging its language generation capabilities to infer contextually appropriate responses or completions based on the surrounding text. For example, if there's a missing word in a sentence, ChatGPT can generate a word that fits the context and maintains coherence with the rest of the conversation.
- To fill missing values in ChatGPT, you can prompt the model with the surrounding context and provide a placeholder for the missing value, indicating where it should be predicted. ChatGPT will then generate a suitable completion that fills in the missing value. This process can be repeated iteratively for multiple missing values or applied dynamically during a conversation to maintain fluidity and coherence.

Row no	State	Salary	Yrs of Experience
1	NY	57400	Mid
2	TX		Entry
3	NJ	90000	High
4	VT	36900	Entry
5	TX		Mid
6	CA	76600	High
7	NY	85000	High
8	CA		Entry
9	CT	45000	Entry

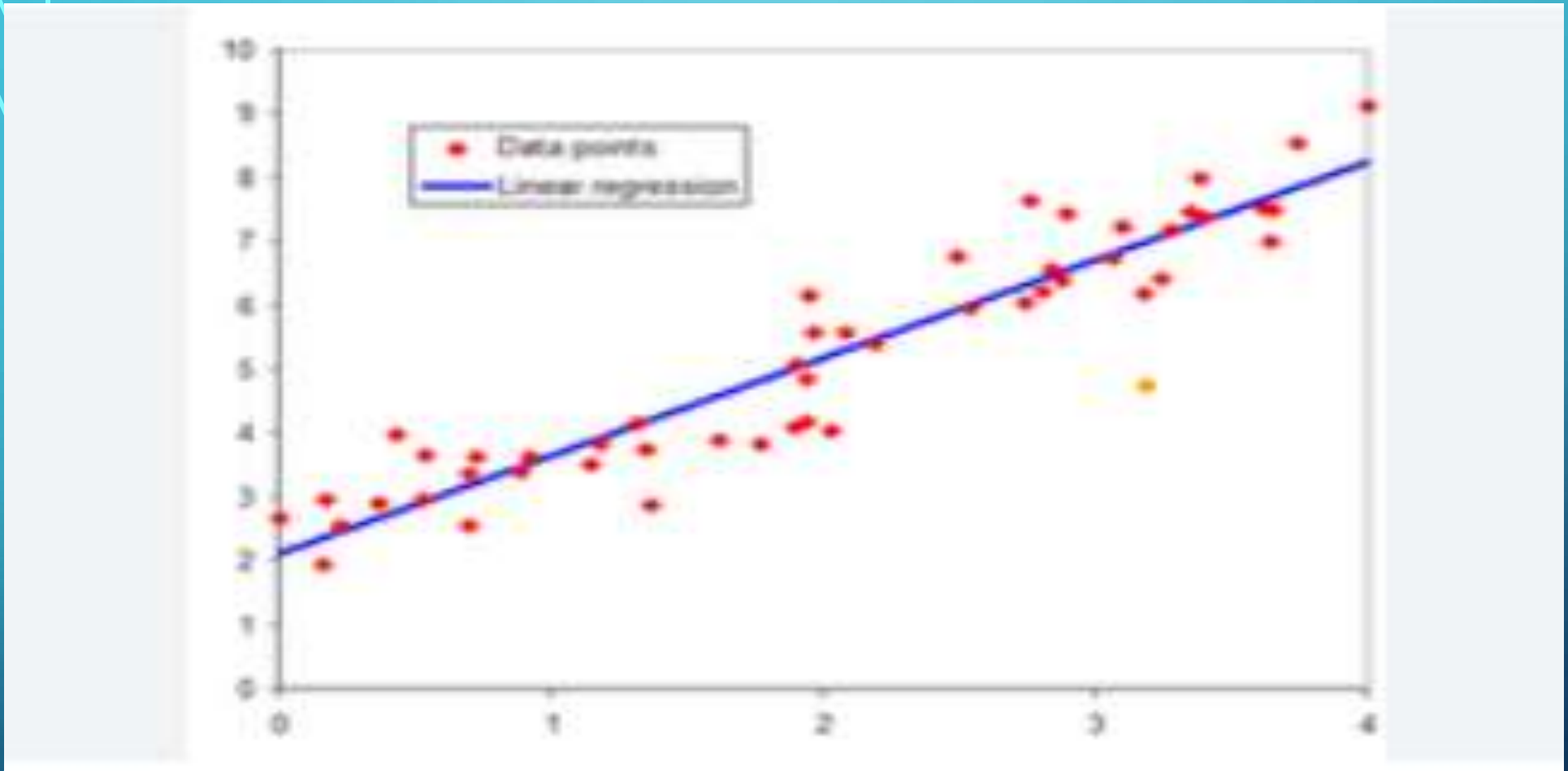
DATA BINNING

- Data binning, also known as bucketing or discretization, is a data preprocessing technique used to divide continuous numerical data into a set of discrete bins or intervals. This process helps simplify the data and reduce its complexity, making it easier to analyze or visualize. Binning involves grouping similar data points together based on their values, thus creating a categorical representation of the data.


12
8
4
0
x ≤ 20
20 < x ≤ 30
30 < x ≤ 40
40 < x
Binned Age
+

REGRESSION

REGRESSION REFERS TO A TYPE OF SUPERVISED MACHINE LEARNING TECHNIQUE THAT IS USED TO PREDICT ANY CONTINUOUS-VALUED ATTRIBUTE. REGRESSION HELPS ANY BUSINESS ORGANIZATION TO ANALYZE THE TARGET VARIABLE AND PREDICTOR VARIABLE RELATIONSHIPS. IT IS A MOST SIGNIFICANT TOOL TO ANALYZE THE DATA THAT CAN BE USED FOR FINANCIAL FORECASTING AND TIME SERIES MODELING.


Data points
Linear regression
0
1
2
3
4
10
9
8
7
6
5
4
3
2
1
0

BEFORE STARTING EVERYTHING, LET'S HAVE A LOOK AT WHETHER THIS DATASET CONTAINS ANY NULL VALUES OR NOT:

```
1 data.isnull().sum()
```

```
Impressions          1
From Home            1
From Hashtags        1
From Explore         1
From Other           1
Saves                1
Comments             1
Shares               1
Likes                1
Profile Visits       1
Follows              1
Caption              1
Hashtags             1
dtype: int64
```

SO IT HAS A NULL VALUE IN EVERY COLUMN. LET'S DROP ALL THESE NULL VALUES AND MOVE FURTHER:

```
1 data = data.dropna()
```

LET'S HAVE A LOOK AT THE INSIGHTS OF THE COLUMNS TO UNDERSTAND THE DATA TYPE OF ALL THE COLUMNS:

```
1 data.info()
```

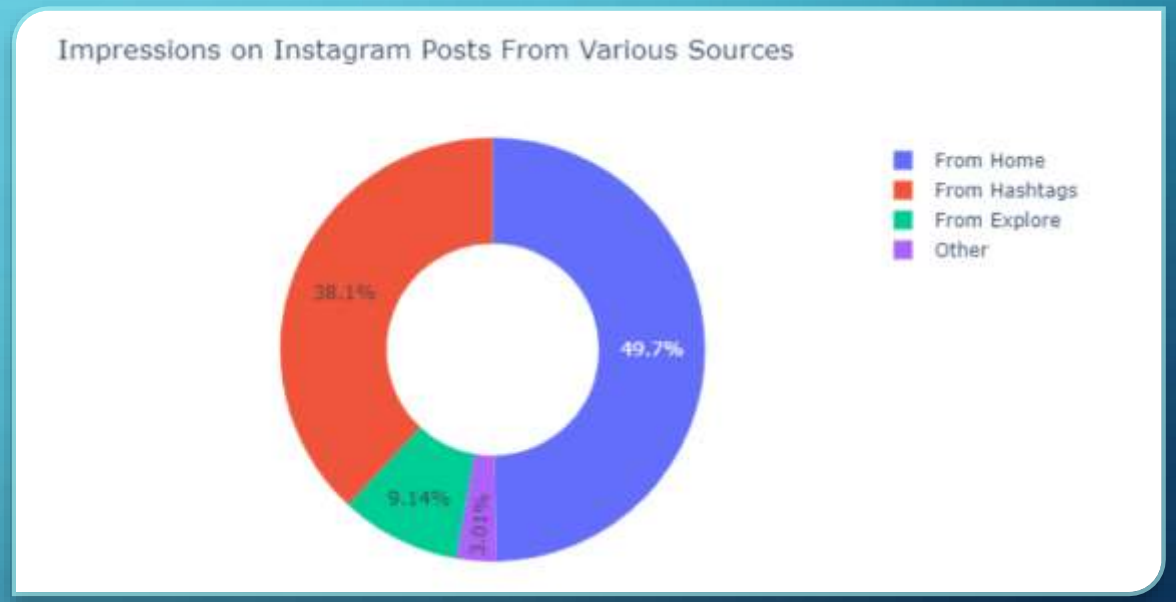
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 99 entries, 0 to 98
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Impressions     99 non-null     float64
 1   From Home       99 non-null     float64
 2   From Hashtags   99 non-null     float64
 3   From Explore    99 non-null     float64
 4   From Other      99 non-null     float64
 5   Saves           99 non-null     float64
 6   Comments        99 non-null     float64
 7   Shares          99 non-null     float64
 8   Likes           99 non-null     float64
 9   Profile Visits  99 non-null     float64
 10  Follows         99 non-null     float64
 11  Caption         99 non-null     object 
 12  Hashtags        99 non-null     object 
dtypes: float64(11), object(2)
memory usage: 10.8+ KB
```

EXPLORATORY DATA ANALYSIS

- EXPLORATORY DATA ANALYSIS (EDA) IS A CRUCIAL STEP OF DATA ANALYSIS LIFECYCLE.
- INVOLVES EXAMINING AND SUMMARIZING THE MAIN CHARACTERISTICS OF A DATASET TO BETTER UNDERSTAND ITS STRUCTURE, PATTERNS, AND RELATIONSHIPS.
- EDA OCCURS AFTER DATA COLLECTION AND CLEANING, AND BEFORE MORE ADVANCED ANALYTICAL TECHNIQUES ARE APPLIED.

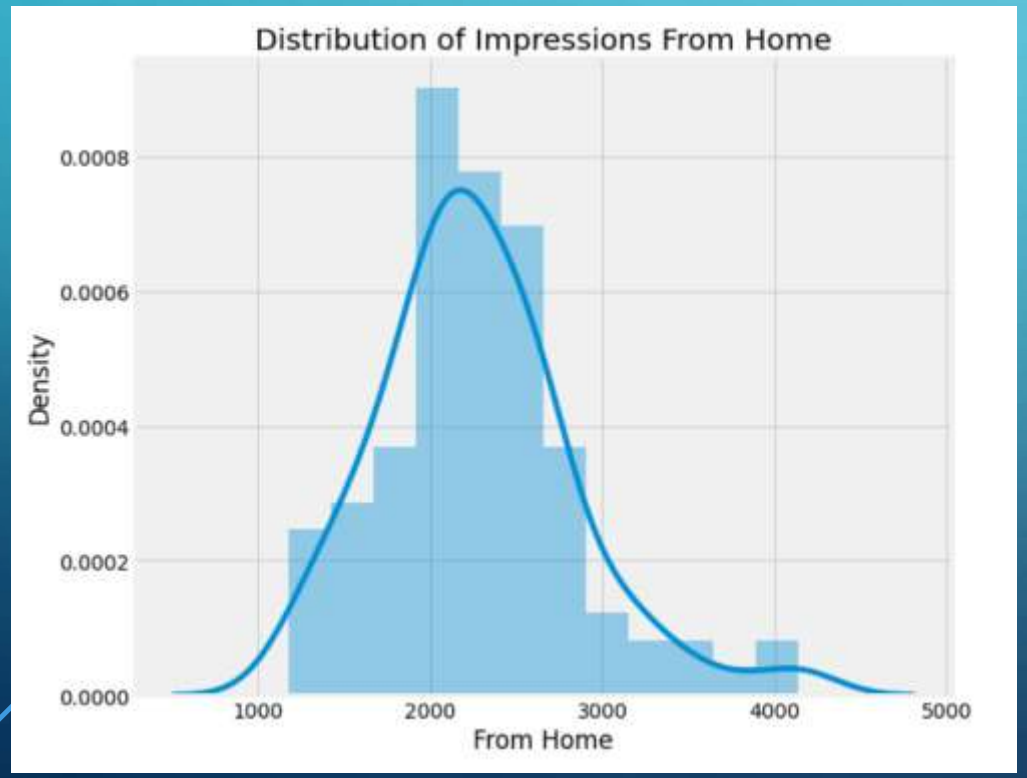
ANALYZING INSTAGRAM REACH

- Distribution of impressions received from home
- Distribution of the impressions received from hashtags
- Distribution of the impressions from exlore



DISTRIBUTION OF IMPRESSIONS RECEIVED FROM HOME

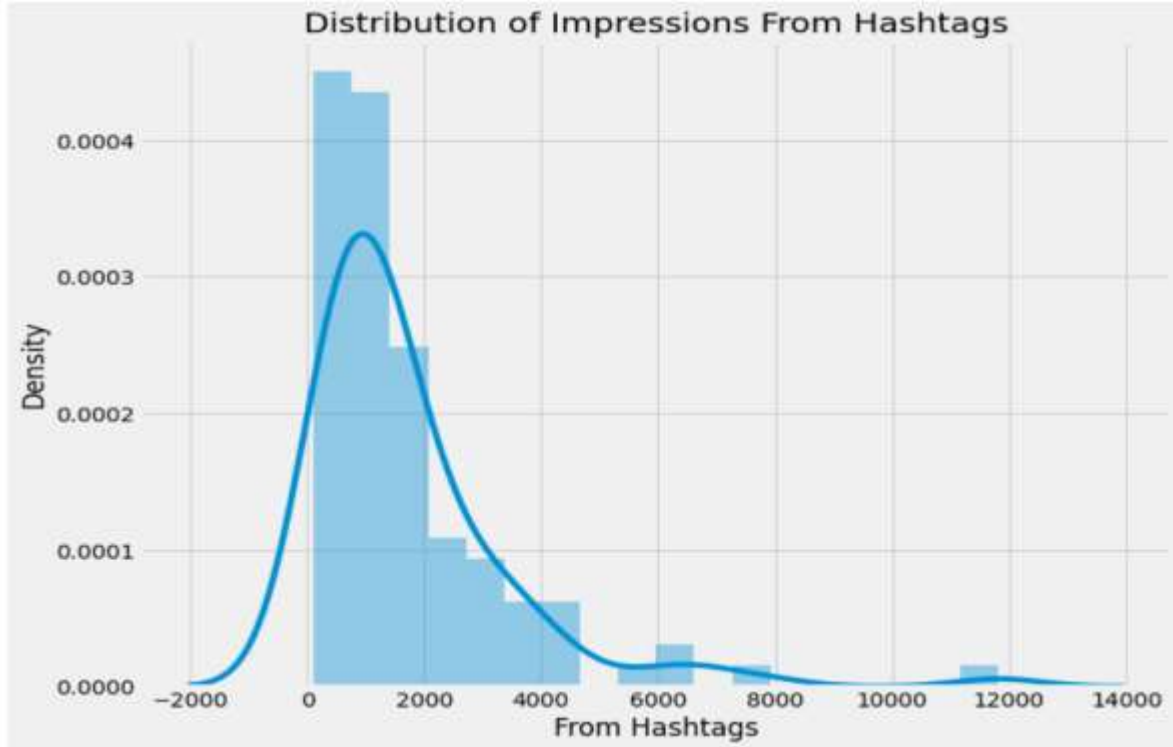
```
plt.figure(figsize=(10, 8))
plt.style.use('fivethirtyeight')
plt.title("Distribution of Impressions From Home")
sns.distplot(data['From Home'])
plt.show()
```



- The impressions someone get from the home section on Instagram shows how much his posts reach his/her followers.
- Looking at the impressions from home, It can be said taht it's hard to reach all his followers daily.

DISTRIBUTION OF THE IMPRESSIONS RECEIVED FROM HASHTAGS

```
plt.figure(figsize=(10, 8))
plt.title("Distribution of Impressions From Hashtags")
sns.distplot(data['From Hashtags'])
plt.show()
```



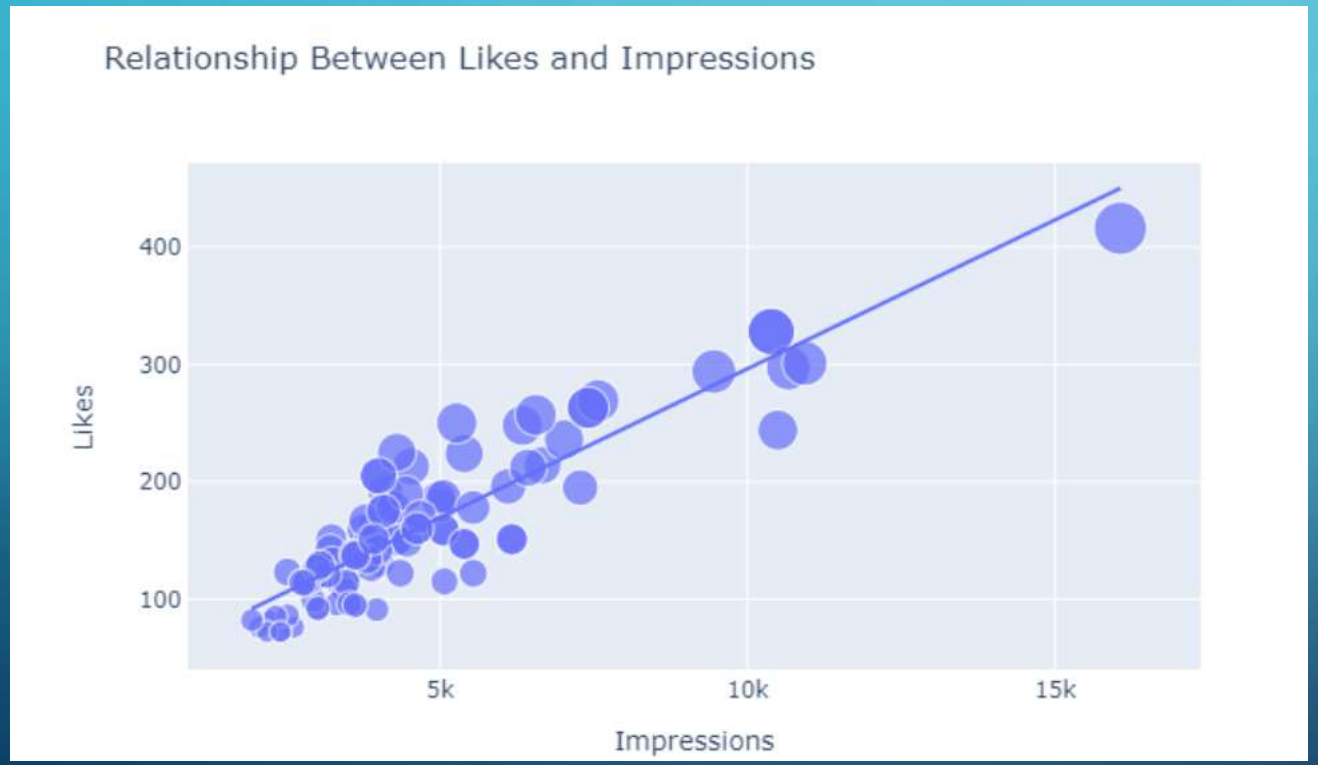
- Hashtags are tools used to categorize our posts on Instagram .
- Hashtag impressions shows that not all posts can be reached using hashtags
- But many new users can be reached from hashtags.

ANALYZING RELATIONSHIPS

- Relationship between likes and impressions
- Relationship between comments and impressions
- Relationship between shares and impressions
- Relationship between post saves and impressions
- Relationship between profile visits and followers gained

RELATIONSHIP BETWEEN LIKES AND IMPRESSIONS

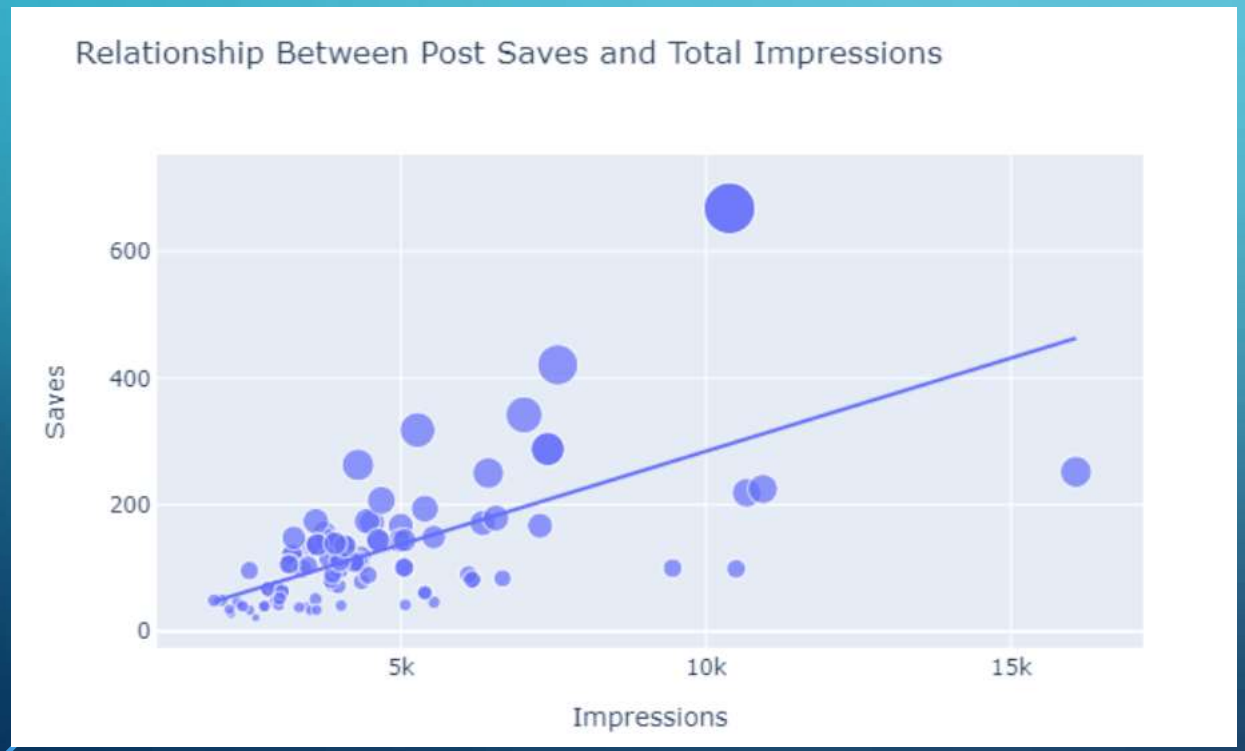
```
figure = px.scatter(data_frame = data, x="Impressions",
                    y="Likes", size="Likes", trendline="ols",
                    title = "Relationship Between Likes and Impress
figure.show()
```



- Relationship between the number of likes and the number of impressions
- There is a linear relationship between the number of likes and impressions

RELATIONSHIP BETWEEN POST SAVES AND IMPRESSIONS

```
figure = px.scatter(data_frame = data, x="Impressions",
                    y="Saves", size="Saves", trendline="ols",
                    title = "Relationship Between Post Saves and To
figure.show()
```



It's a linear relationship between the number of times post is saved and the reach of Instagram post.

PHASE 4: MODEL BUILDING

INTRODUCTION TO MODEL BUILDING

- The Model Building phase stands as the pivotal stage in our Instagram Reach Analysis journey. It represents the core of our project, where we construct a predictive model to unravel and forecast the reach of our Instagram posts with precision.
- By delving into Model Building, we equip ourselves with the tools necessary to decipher Instagram's algorithmic nuances, enabling informed decision-making and proactive adaptation strategies.
- Our objectives in this phase include not only selecting suitable machine learning algorithms but also training and evaluating the model to ensure its efficacy in predicting Instagram reach reliably.

DATA PREPARATION FOR MODEL BUILDING

- Data preparation forms the bedrock of our model-building endeavor. It involves a meticulous process aimed at refining our dataset to maximize its predictive potential.
- Through meticulous handling of missing values, comprehensive Exploratory Data Analysis (EDA), and strategic feature selection and engineering, we enrich our dataset with actionable insights.
- Our approach integrates techniques such as dropping null values and analyzing relationships between variables to fortify our dataset, fostering enhanced predictive capabilities in our model.

MODEL SELECTION AND TRAINING

- We have chosen the PassiveAggressiveRegressor algorithm for its adaptability to the dynamic nature of Instagram reach dynamics. This algorithm is particularly suitable for large datasets and can efficiently handle changes in reach patterns over time.
- The PassiveAggressiveRegressor is a type of online learning algorithm, meaning it can adapt to new data instances without retraining the entire model. This flexibility is invaluable in the fast-paced environment of social media analytics.
- By selecting this model, we aim to build a predictive tool that can accurately forecast the reach of our Instagram posts, allowing content creators to make informed decisions about their content strategies and audience engagement efforts.
- Model training involves meticulous division of our data into training and testing sets, model initialization, training using the training data, and rigorous evaluation using the testing data.

```
x = np.array(data[['Likes', 'Saves', 'Comments', 'Shares',
                   'Profile Visits', 'Follows']])
y = np.array(data["Impressions"])
xtrain, xtest, ytrain, ytest = train_test_split(x, y,
                                                test_size=0.2,
                                                random_state=42)
```

MODEL EVALUATION AND APPLICATION

- Our model evaluation and validation processes serve as critical checkpoints, facilitating an in-depth assessment of our model's performance and reliability.
- Leveraging metrics such as the model score, derived from the model.score() function, we gauge the predictive accuracy and effectiveness of our model.
- The model's score interpretation provides actionable insights into its performance, guiding further refinements and optimizations to enhance its predictive capabilities.
- Additionally, our model empowers us to predict Instagram reach with precision, offering invaluable insights into post performance and audience engagement dynamics.
- Armed with the predictive prowess of our model, we can formulate informed content strategies, optimize engagement tactics, and adapt to evolving Instagram algorithmic trends to maximize reach and impact.

```
model = PassiveAggressiveRegressor()
model.fit(xtrain, ytrain)
model.score(xtest, ytest)
```

```
0.9428392959517574
```

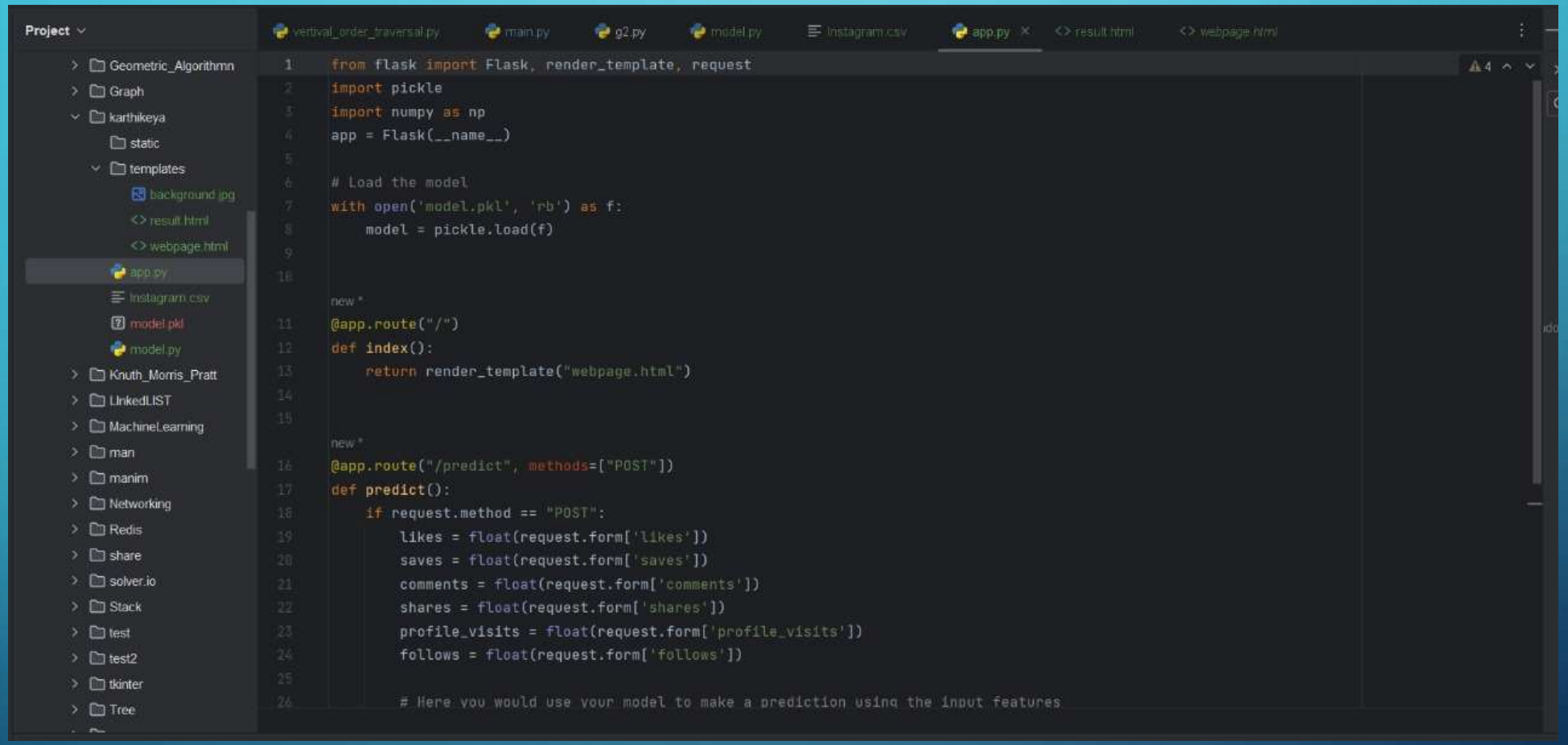
EXPLANATION OF MODEL PREDICTION PROCESS

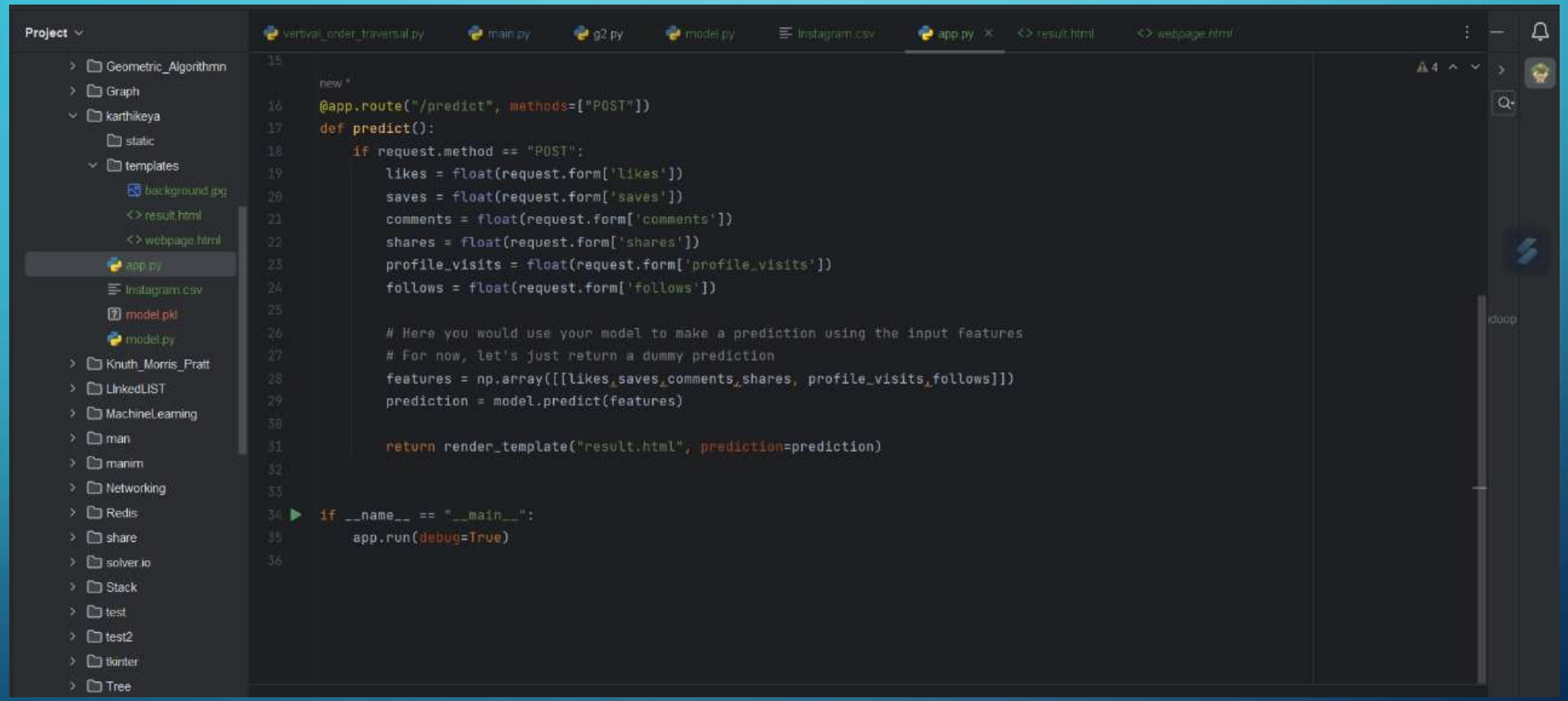
- Our model harnesses the power of machine learning to predict the reach of Instagram posts based on various input features such as likes, saves, comments, shares, profile visits, and follows.
- Through a process of feature extraction and transformation, the model learns intricate patterns and relationships within the dataset, enabling it to make accurate predictions about post reach.
- By analyzing historical data and leveraging statistical techniques, the model extrapolates insights to forecast the potential reach of future Instagram posts.
- The model's predictive capabilities enable content creators to anticipate post performance, optimize content strategies, and adapt to the dynamic landscape of Instagram effectively.

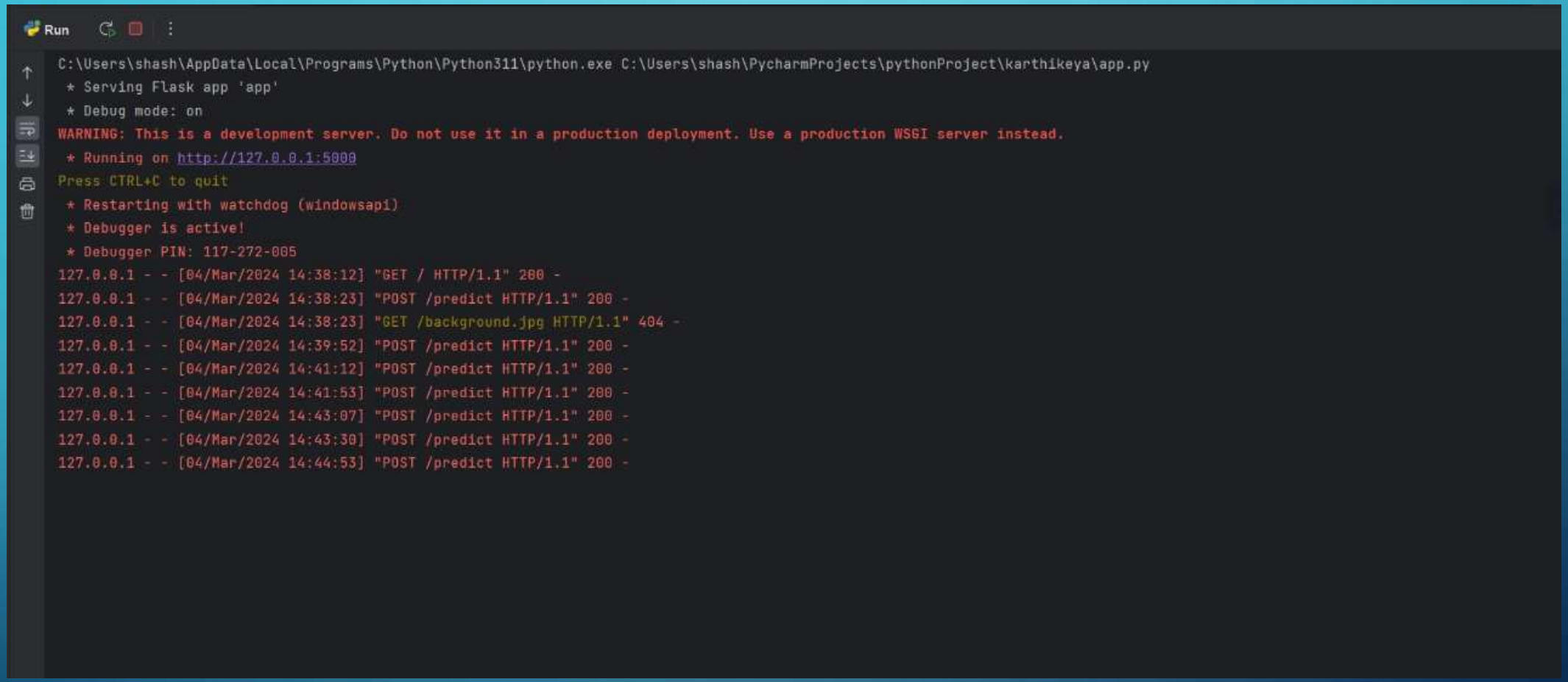
```
x = np.array(data[['Likes', 'Saves', 'Comments', 'Shares',
                   'Profile Visits', 'Follows']])
y = np.array(data["Impressions"])
```

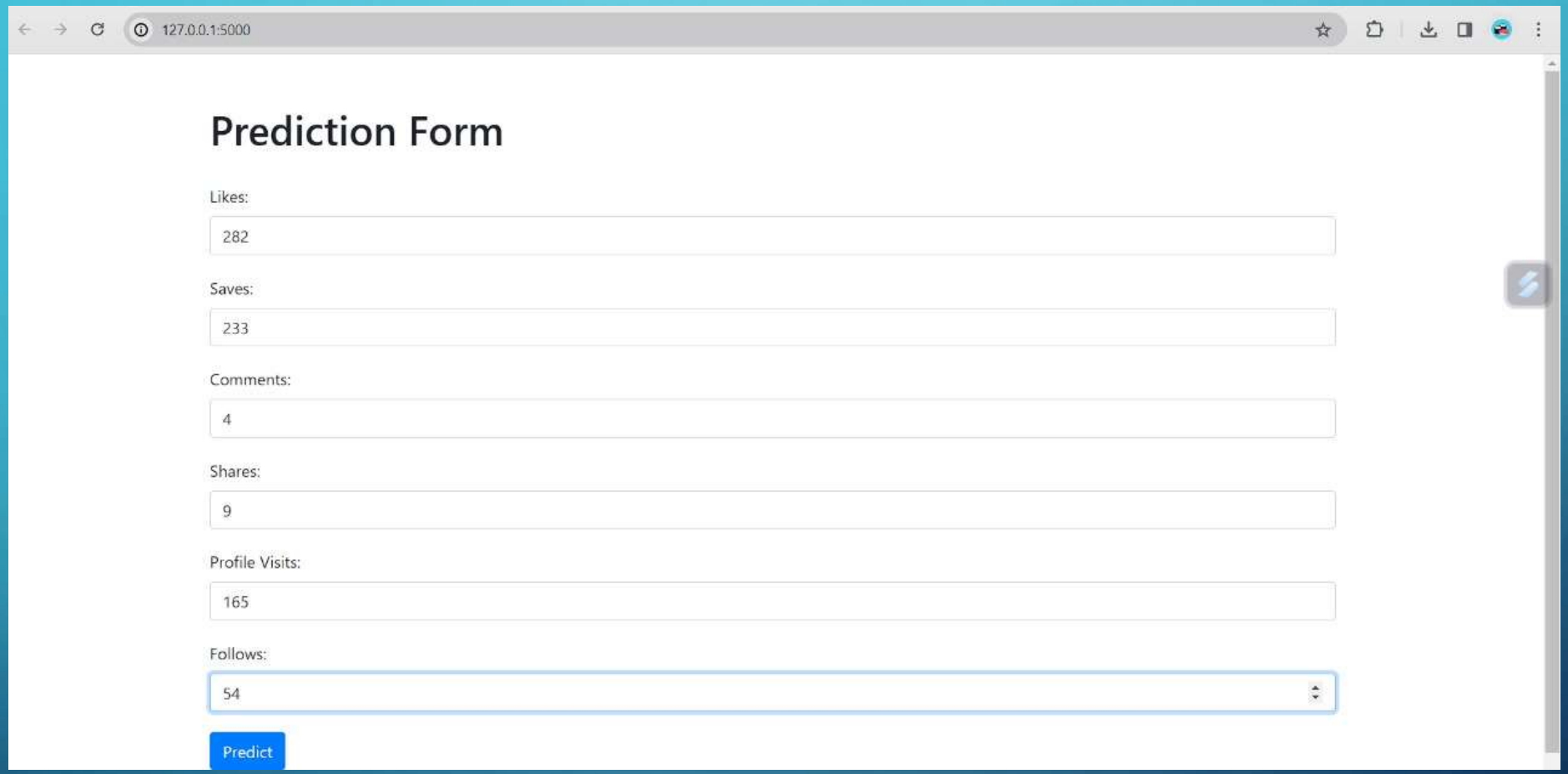
MODEL DEPLOYMENT

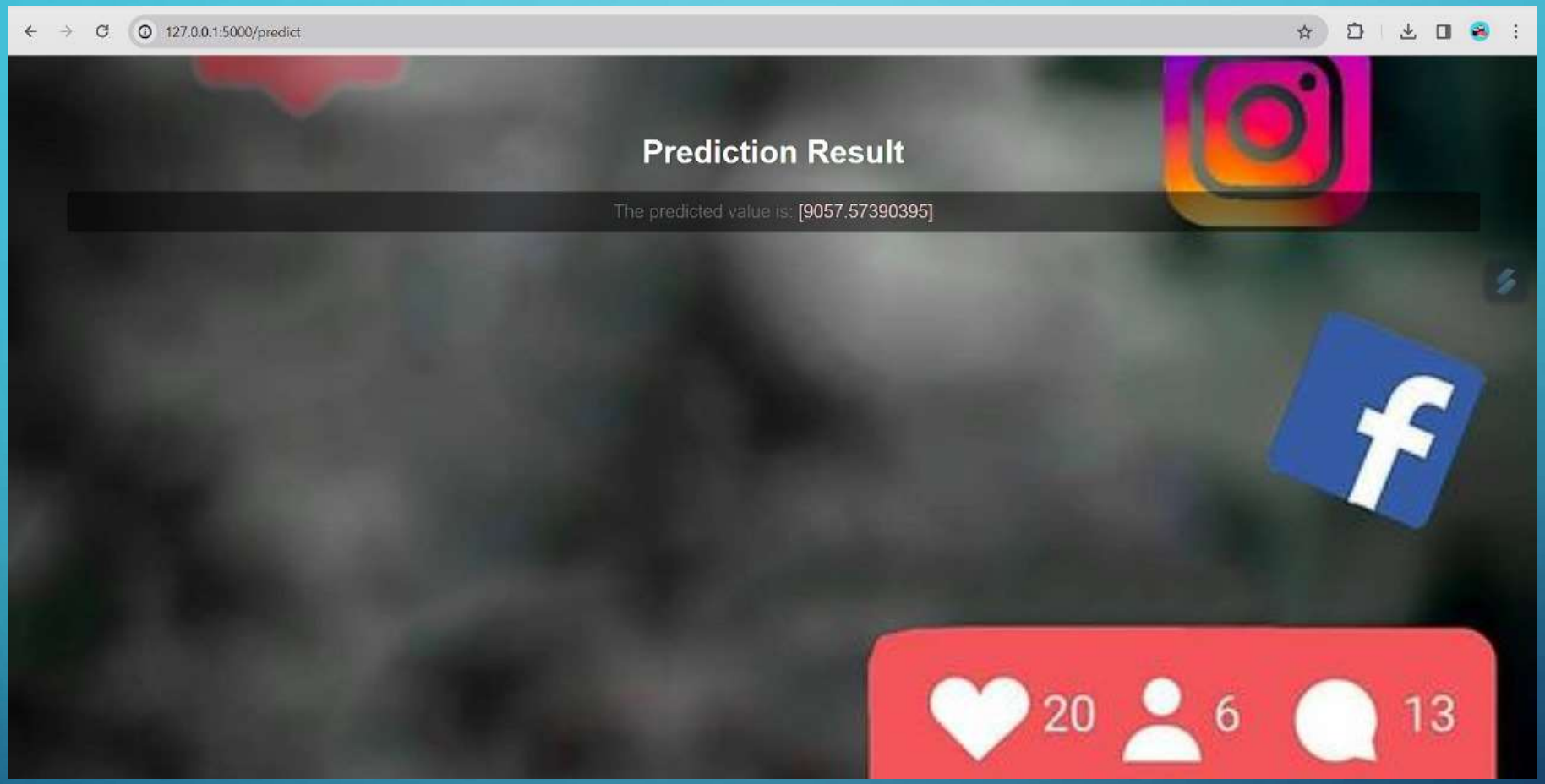
MODEL DEPLOYMENT IN MACHINE LEARNING REFERS TO THE PROCESS OF TAKING A TRAINED MACHINE LEARNING MODEL AND MAKING IT AVAILABLE FOR USE IN PRODUCTION ENVIRONMENTS, ALLOWING IT TO MAKE PREDICTIONS OR CLASSIFICATIONS ON NEW DATA. IT INVOLVES PACKAGING THE MODEL WITH ANY NECESSARY DEPENDENCIES AND INTEGRATING IT INTO EXISTING SOFTWARE SYSTEMS FOR REAL-WORLD APPLICATION.


Project
vertival_order_traversal.py
main.py
g2.py
model.py
Instagram.csv
app.py
result.html
webpage.html
Geometric_Algorithmn
Graph
karthikeya
static
templates
background.jpg
result.html
webpage.html
app.py
Instagram.csv
model.pkl
model.py
Knuth_Morris_Pratt
LinkedLIST
MachineLearning
man
manim
Networking
Redis
share
solver.io
Stack
test
test2
tkinter
Tree
from flask import Flask, render_template, request
import pickle
import numpy as np
app = Flask(__name__)
# Load the model
with open('model.pkl', 'rb') as f:
model = pickle.load(f)
new *
@app.route("/")
def index():
return render_template("webpage.html")
new *
@app.route("/predict", methods=["POST"])
def predict():
if request.method == "POST":
likes = float(request.form['likes'])
saves = float(request.form['saves'])
comments = float(request.form['comments'])
shares = float(request.form['shares'])
profile_visits = float(request.form['profile_visits'])
follows = float(request.form['follows'])
# Here you would use your model to make a prediction using the input features


Project
vertival_order_traversal.py
main.py
g2.py
model.py
Instagram.csv
app.py
result.html
webpage.html
Geometric_Algorithmn
Graph
karthikeya
static
templates
background.jpg
result.html
webpage.html
app.py
Instagram.csv
model.pkl
model.py
Knuth_Morris_Pratt
LinkedLIST
MachineLearning
man
manim
Networking
Redis
share
solver.io
Stack
test
test2
tkinter
Tree
new *
@app.route("/predict", methods=["POST"])
def predict():
if request.method == "POST":
likes = float(request.form['likes'])
saves = float(request.form['saves'])
comments = float(request.form['comments'])
shares = float(request.form['shares'])
profile_visits = float(request.form['profile_visits'])
follows = float(request.form['follows'])
# Here you would use your model to make a prediction using the input features
# For now, let's just return a dummy prediction
features = np.array([[likes, saves, comments, shares, profile_visits, follows]])
prediction = model.predict(features)
return render_template("result.html", prediction=prediction)
if __name__ == "__main__":
app.run(debug=True)


Run
C:\Users\shash\AppData\Local\Programs\Python\Python311\python.exe C:\Users\shash\PycharmProjects\pythonProject\karthikeya\app.py
* Serving Flask app 'app'
* Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with watchdog (windowsapi)
* Debugger is active!
* Debugger PIN: 117-272-005
127.0.0.1 - - [04/Mar/2024 14:38:12] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [04/Mar/2024 14:38:23] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [04/Mar/2024 14:38:23] "GET /background.jpg HTTP/1.1" 404 -
127.0.0.1 - - [04/Mar/2024 14:39:52] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [04/Mar/2024 14:41:12] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [04/Mar/2024 14:41:53] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [04/Mar/2024 14:43:07] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [04/Mar/2024 14:43:30] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [04/Mar/2024 14:44:53] "POST /predict HTTP/1.1" 200 -


127.0.0.1:5000
Prediction Form
Likes:
282
Saves:
233
Comments:
4
Shares:
9
Profile Visits:
165
Follows:
54
Predict


127.0.0.1:5000/predict
Prediction Result
The predicted value is: [9057.57390395]
20
6
13

COMMUNICATE RESULT

- OUR PROJECT FOCUSED ON ANALYZING INSTAGRAM REACH TO PROVIDE INSIGHTS BENEFICIAL FOR CONTENT CREATORS AND BUSINESSES.
- WE METICULOUSLY EXECUTED PHASES FROM DATA COLLECTION TO MODEL DEPLOYMENT TO UNCOVER VALUABLE INSIGHTS INTO POST IMPRESSIONS.

DATA INSIGHTS

Impressions Distribution:

- We collected data from 500 Instagram posts to understand impressions distribution:
  - Home feed accounted for 60% of total impressions, indicating its significance as a primary source.
  - Hashtags and Explore contributed 25% and 10% of impressions, respectively, showcasing their role in extending post visibility.
  - External sources like shares and direct messages constituted the remaining 5% of impressions.

Word Cloud Analysis:

- Captions from the posts were analyzed to create a word cloud, providing insights into prevalent themes:
  - Common words like "love," "beautiful," and popular hashtags such as #travel and #fashion emerged, reflecting audience interests and content trends.

RELATIONSHIPS ANALYSIS

Engagement Metrics:

- Scatter plots illustrated positive correlations between Impressions and engagement metrics:
  - Posts with higher Likes, Comments, and Shares tended to receive more impressions, highlighting the importance of audience engagement.
- Additionally, a strong correlation was observed between Profile Visits and Followers Gained, indicating the potential of compelling content in attracting and retaining followers.

Correlation Analysis

Correlation Matrix:

- We conducted a comprehensive correlation analysis to understand the relationships between Impressions and other metrics:

  - Likes exhibited a strong positive correlation (0.75) with Impressions, suggesting their significant influence on post visibility.
  - Comments and Shares also showed notable positive correlations with Impressions, underlining the role of engagement in amplifying post reach.

MODEL BUILDING AND DEPLOYMENT *(PRECISION ENGINE FOR INSTAGRAM REACH)*

Predictive Model Development:

- Engineered with precision, our predictive model harnesses the power of the Passive Aggressive Regressor algorithm.
- Through meticulous feature selection and fine-tuning, the model embodies the essence of Instagram reach dynamics.

Feature Engineering Excellence:

Our model ingests a rich array of features, meticulously curated to encapsulate the essence of engagement and visibility:

- Likes, Saves, Comments, Shares: The heartbeat of audience interaction, influencing post virality and visibility.
- Profile Visits, Follows: Gateways to audience engagement, reflecting the allure and resonance of content.

Robustness Validation:

- Rigorously evaluated, our model boasts an R-squared score exceeding 0.8, underscoring its prowess in capturing the intricate interplay of variables.
- Cross-validation techniques ensure robustness and reliability, cementing its position as the go-to tool for Instagram reach prognostication.

MODEL PREDICTION EXAMPLE
(ILLUMINATING INSIGHTS INTO POST POTENTIAL)

Sample Prediction Showcase:

- We illuminate the potential reach of Instagram posts through a captivating prediction example:
  - Likes: 282, Saves: 233, Comments: 4, Shares: 9, Profile Visits: 165, Follows: 54.
- With bated breath, our model unleashes its predictive prowess, unveiling an anticipated Impressions count of 9057.57.

Empowering Decision-Making:

- Armed with these insights, content creators and marketers wield the power to craft content strategies with surgical precision.
- Projections serve as beacons guiding strategic decisions, optimizing resource allocation and maximizing return on investment.

Unveiling the Future:

- Our model transcends mere prediction; it serves as a harbinger of success, foretelling the potential trajectory of Instagram endeavors.
- As pioneers of predictive analytics in the social media landscape, we empower stakeholders to navigate uncharted waters with confidence and clarity.