

Why ML?



Let's apply for a
internship/job
position

- ★ Parul from ISP Team Inbox  Invitation: Pranav from K. K. Wagh Institute Of Engineering Education And Research represent
- ★ ISP Team from Inter. Inbox  Congratulations, Pranav! You're selected to apply to ISP Program - Internshala Trainings Internshala
- ★ ISP Team from Inter. Inbox  Earn ₹7 lacs for your bank account NOW - Internshala Trainings Deal expires soon! [URLTEXT]
- ★ ISP Team from Inter. Inbox  Pranav's iPhone is out for delivery - Internshala Trainings Internshala Student Partner Program Get
- ★ ISP Team from Inter. Inbox  Only K. K. Wagh Institute Of Engineering Education And Research students should open this email
- ★ ISP Team from Inter. Inbox  K. K. Wagh Institute Of Engineering Education And Research students pocket money credited -
- ★ ISP Team from Inter. Inbox  Pranav, represent us in K. K. Wagh Institute Of Engineering Education And Research - Internshala
- ★ Internshala Inbox  Pranav, 6 tips to handle career rejections! - Internshala (Scholiverse Educare Pvt. Ltd.) 901 A/B, Iri
- ★ ISP Team from Inter. Inbox  Pranav, FINAL 3 hours left | Chance to win iPhone 14! - Internshala Trainings Internshala Student Pa
- ★ Parul from ISP Team Inbox  iPhone 14 sailing away, Pranav | FINAL few hours left! - Internshala Trainings Internshala Student Pa
- ★ Parul from ISP Team Inbox  Pranav, Learning mangoge, iPhone 14 bhi denge! | Application deadline extended - Internshala Trainings Internshala Student Pa
- ★ ISP Team from Inter. Inbox  Pranav, we know what you did! - Internshala Trainings Internshala Student Partner program You igno
- ★ ISP Team from Inter. Inbox  Pranav, INR 7 Lacs for 1% equity! Deal done karein? - Internshala Trainings Deal expires soon! [URLTEXT]

Building your
own email
service

Email Composition and Sending
Email Reception and Storage
Inbox Management
Security Features
Attachments Handling
Contact Management

Spam and Junk Mail Filtering

```
email_text = "For you: Win a free iPhone! Just click here 4 ur chance 2 claim now!"  
  
def is_spam(email: str) -> bool:  
    spam_keywords = ["free", "For you", "claim now"]  
    for word in spam_keywords:  
        if word in email:  
            return True  
    return False  
  
# Test the function  
print(is_spam(email_text)) # Should print True
```



```
email_text = "Special offer 4 u: Get a $500 voucher! Don't miss out, claim urs  
today!"
```

```
def is_spam(email: str) -> bool:  
    spam_keywords = ["free", "4 u", "claim urs"]  
    for word in spam_keywords:  
        if word in email:  
            return True  
    return False
```

```
# Test the function  
print(is_spam(email_text)) # Should print True
```

Long List of Rules

```
● ● ●

# List of common spam words
spam_words = ["Congratulations", "You've", "won", "prize", "Click", "here",
"claim", "reward", "now", "free", "gift", "voucher", "offer"]

# Function to generate a spam message based on specific spam words
def generate_spam_message():
    message = []
    if "Congratulations" in spam_words:
        message.append("Congratulations")
    if "You've" in spam_words:
        message.append("You've")
    if "won" in spam_words:
        message.append("won")
    if "a" in spam_words:
        message.append("a")
    if "prize" in spam_words:
        message.append("prize")
    if "Click" in spam_words:
        message.append("Click")
    if "here" in spam_words:
        message.append("here")
    if "to" in spam_words:
        message.append("to")
```

Long List of Rules

Complex Problems

Fluctuating Environments

Understand large amounts of data.

Recommend

Reel Recommendation

Movie Recommendation

Product Recommendation

Risk Analysis

Business Analysis

Understanding Machine Learning Beyond HYPE

By: Pranav Ahire

Talk Flow

Data Manipulation
ML Project Workflow
Maths
DL, LLM's & Gen AI

What is Machine Learning?

science of programming computers so they can learn from **data**.

field of study that gives computers the ability to learn without being **explicitly programmed**

A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with **experience E**.

Data Model

**What is
Data?**

• • •

Customer_ID: 2002, Transaction_ID: 1002, Product_ID: 30002, Shipping_Address: 456 Oak St, Payment_Method: PayPal, Transaction_Date: 2023-07-22, Country: USA, Price: 9.99, Quantity: 1

Customer_ID: 2001, Transaction_ID: 1003, Transaction_Date: 2023-07-23, Country: USA, Customer_ID: 2001, Payment_Method: Credit Card, Shipping_Address: 123 Elm St, Apt 4, Product_ID: 30001, Quantity: 3

Customer_ID: 2003, Product_ID: 30004, Quantity: 5, Transaction_Date: 2023-07-24, Customer_ID: 2003, Product_ID: 30004, Price: 29.99, Payment_Method: Debit Card, Shipping_Address: 789 Pine St, Country: USA

Customer_ID: 2001, Shipping_Address: 123 Elm St, Apt 4, Payment_Method: Credit Card, Customer_ID: 2001, Product_ID: 30001, Price: 19.99, Transaction_Date: 2023-07-21, Transaction_ID: 1001, Quantity: 2

Customer_ID: 2004, Transaction_Date: 2023-07-25, Payment_Method: Credit Card, Shipping_Address: 987 Maple St, Country: USA, Transaction_ID: 1005, Product_ID: 30004, Customer_ID: 2004, Price: 15.99



Transaction_ID	Customer_ID	Transaction_Date	Product_ID	Quantity	Price
1001	2001	2023-07-21	30001	2	19.99
1002	2002	2023-07-22	30002	1	9.99
1003	2003	2023-07-23	30001	3	19.99
1004	2003	2023-07-24	30003	5	4.99
1005	2004	2023-07-25	30004	2	15.99



Data Engineer identifies the business problem and translates these to data services and engineering outcomes

Information that has been translated into a form that is efficient for movement or processing.

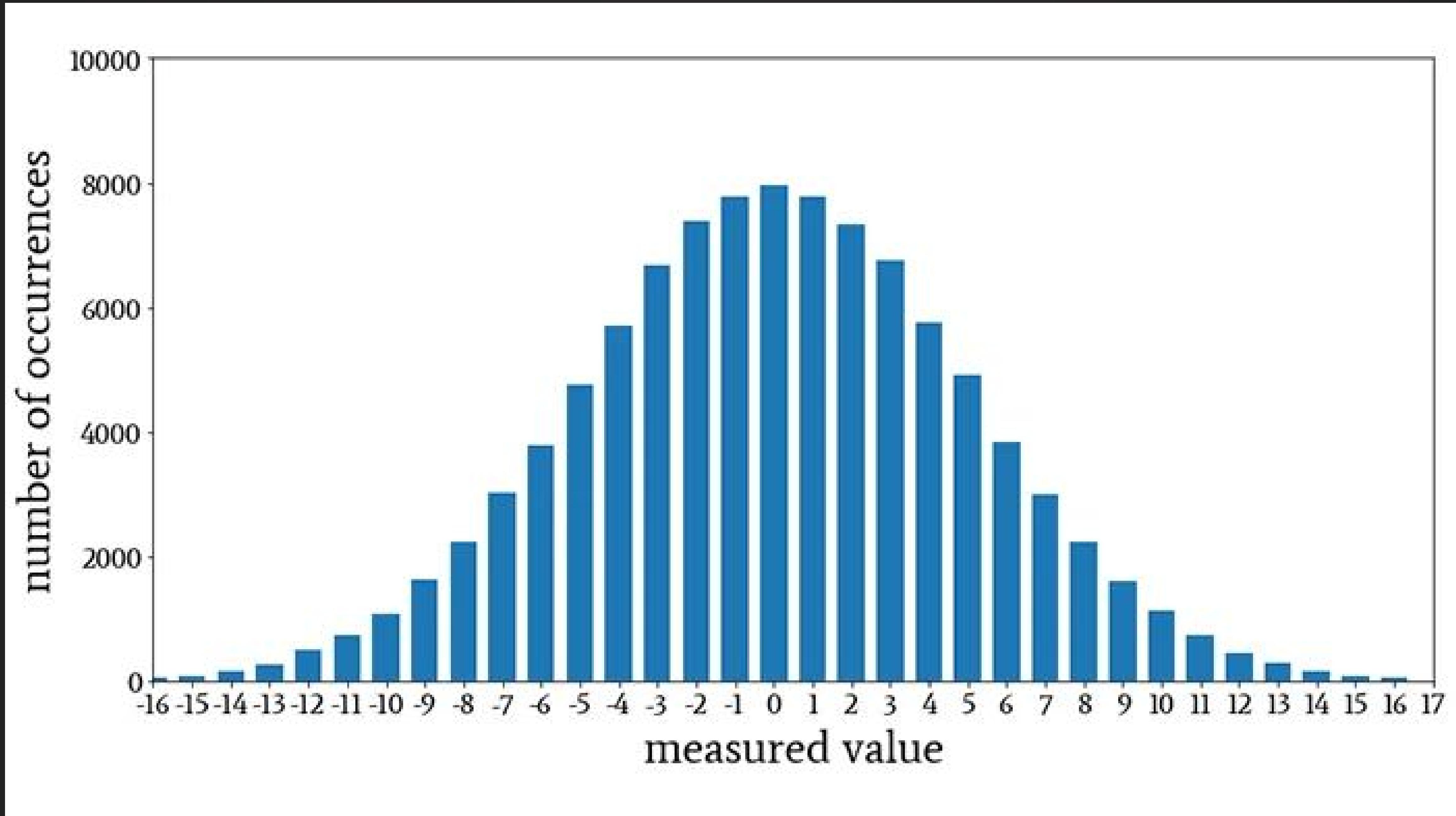
How Data is Generated?

**Random (Unbiased
Sample)**

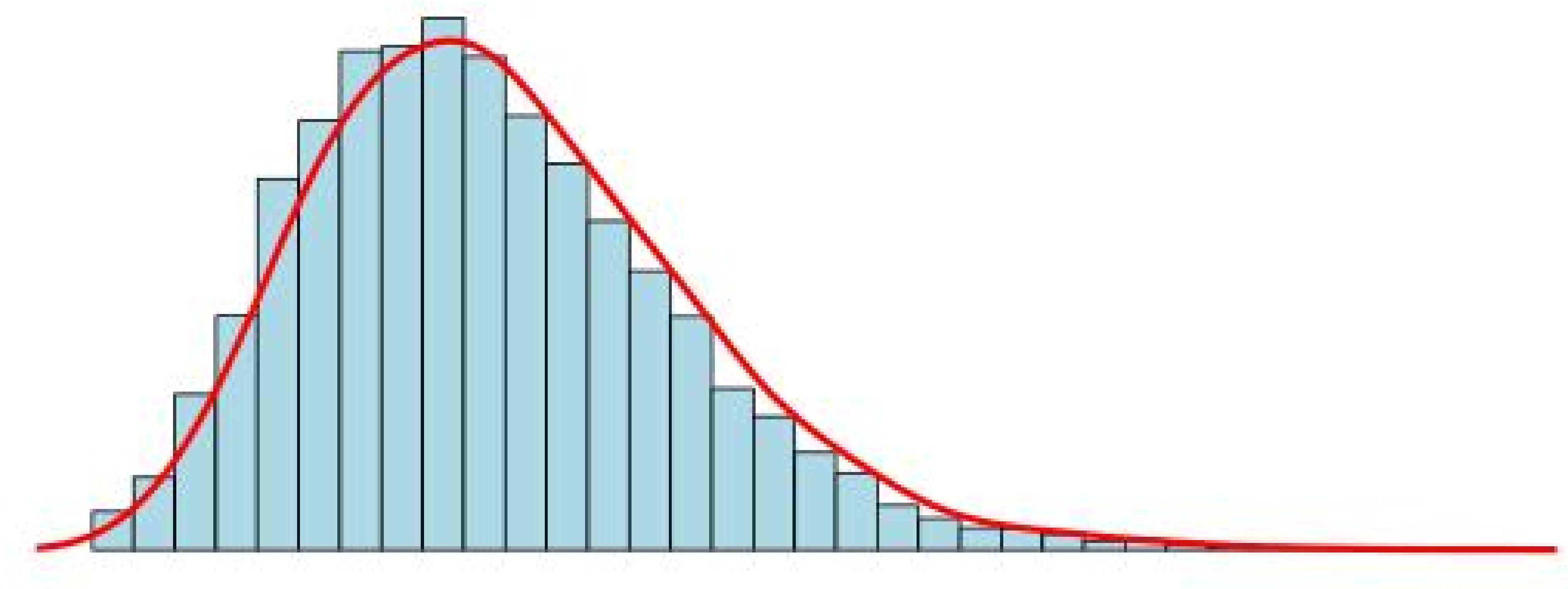
Population

understanding
Data

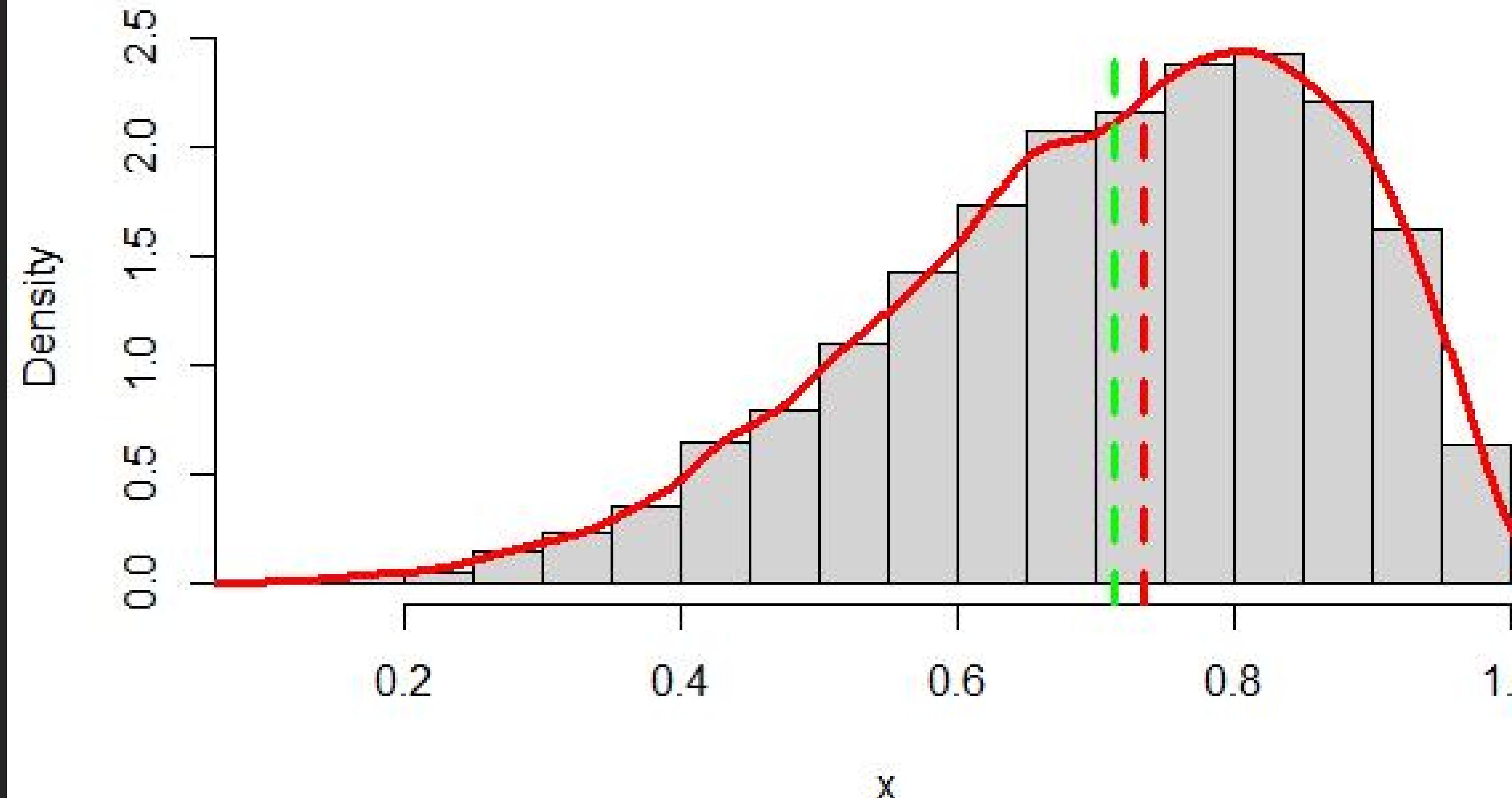
Normal Distributions

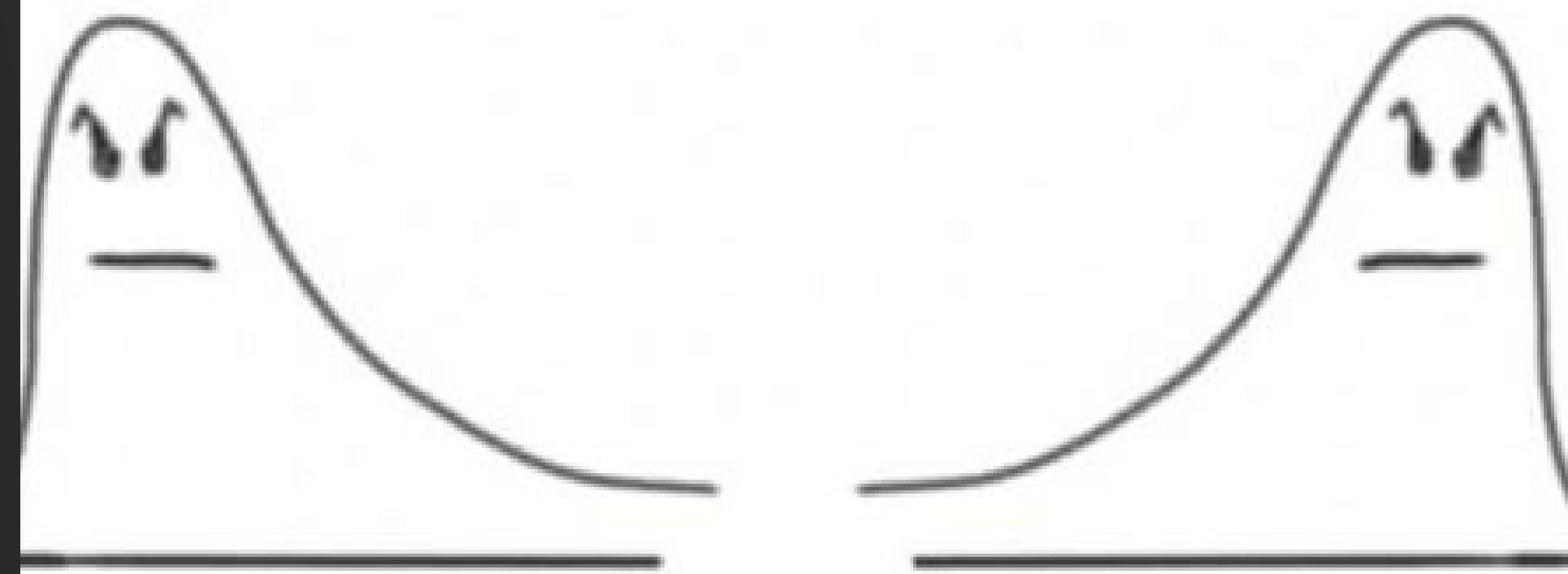
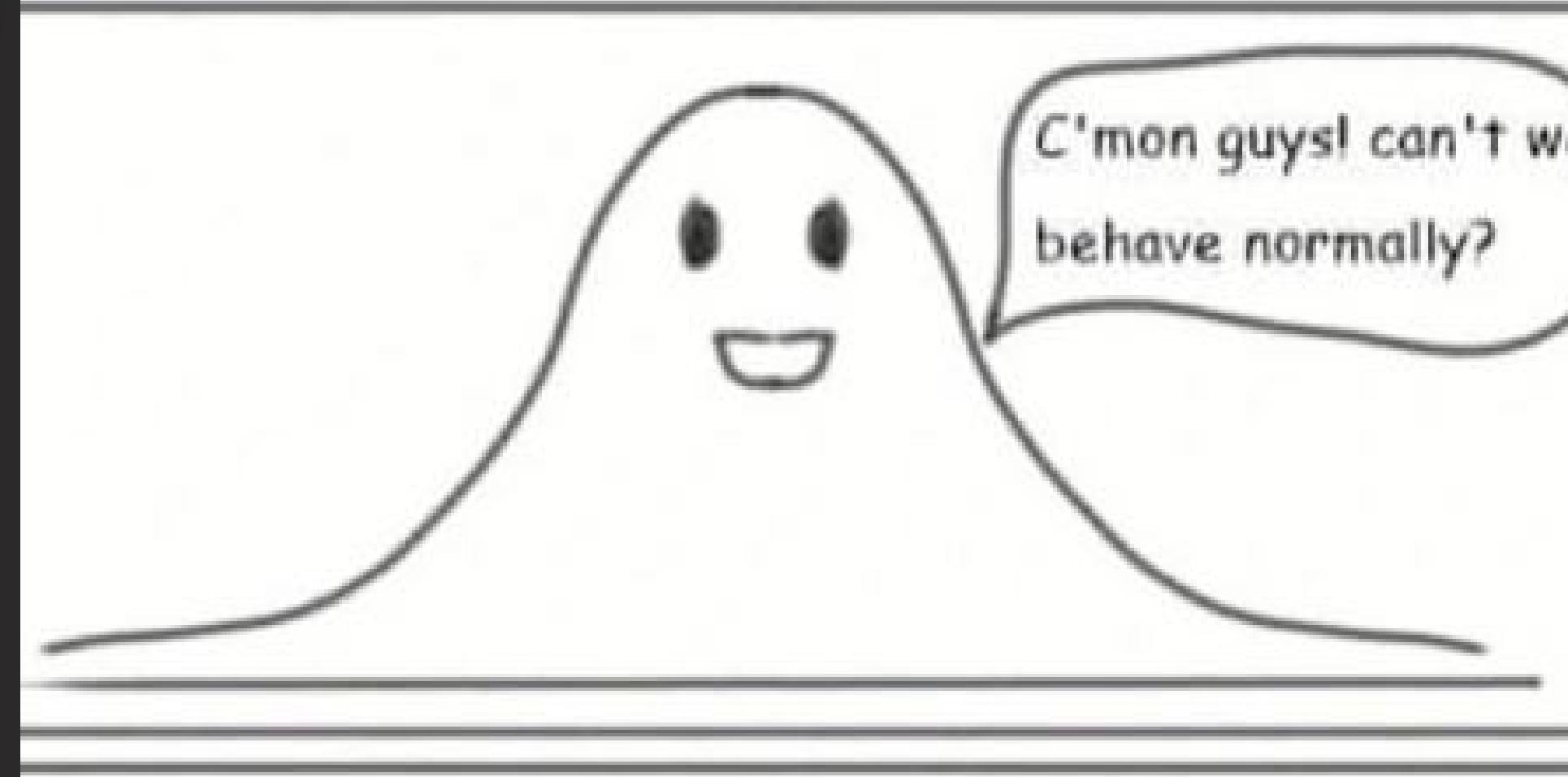
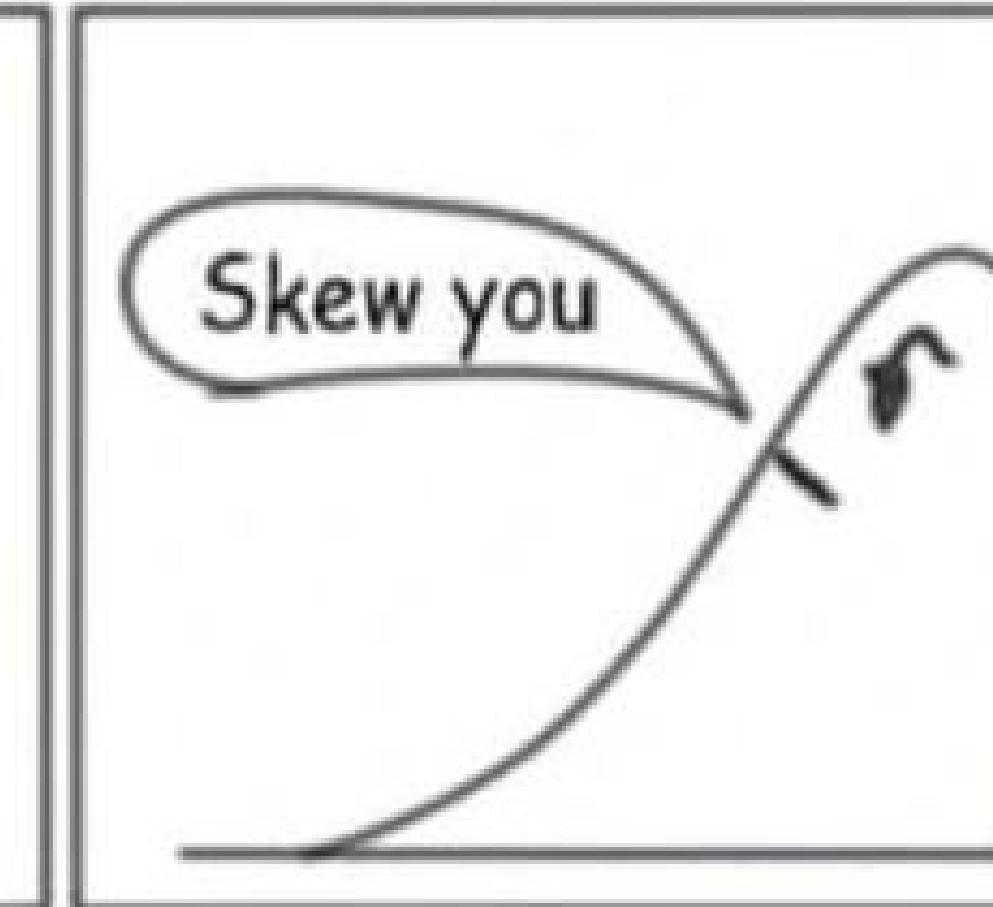
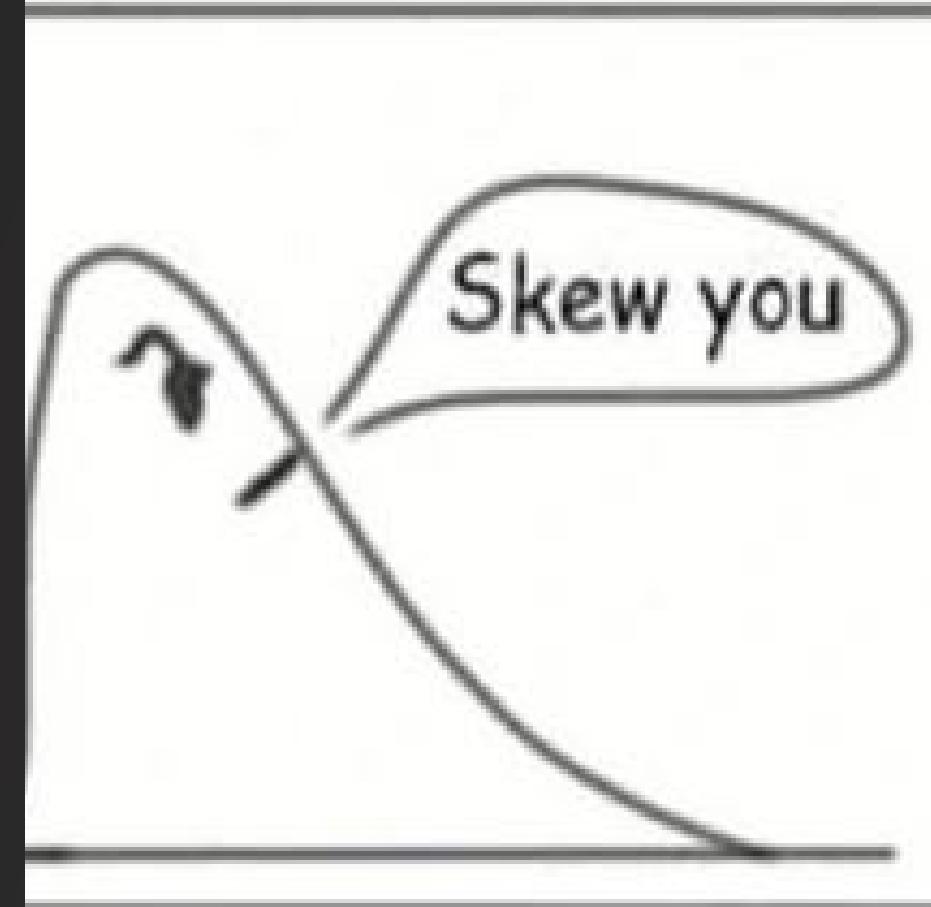


Positive Skewed



Negative or Left Skewness



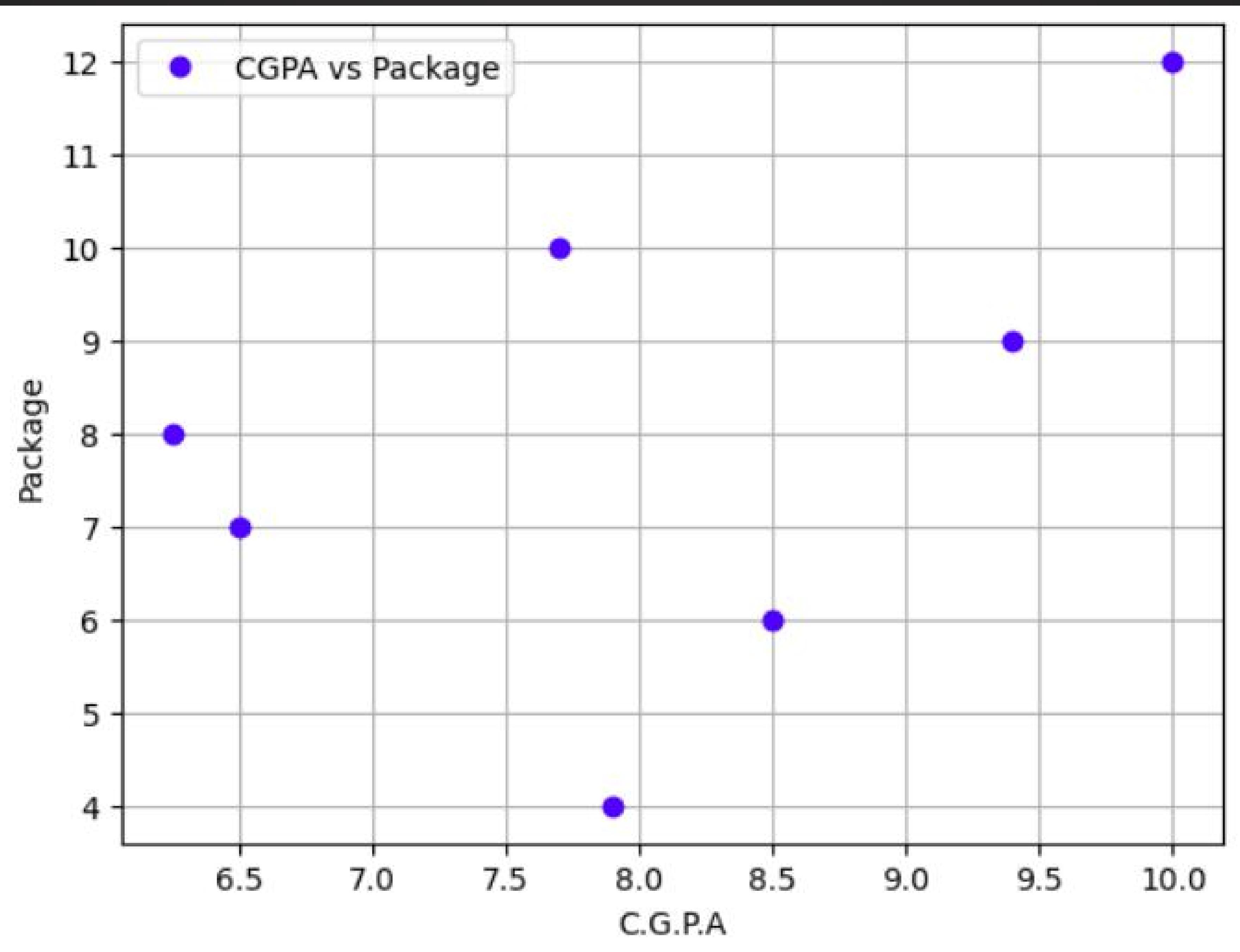


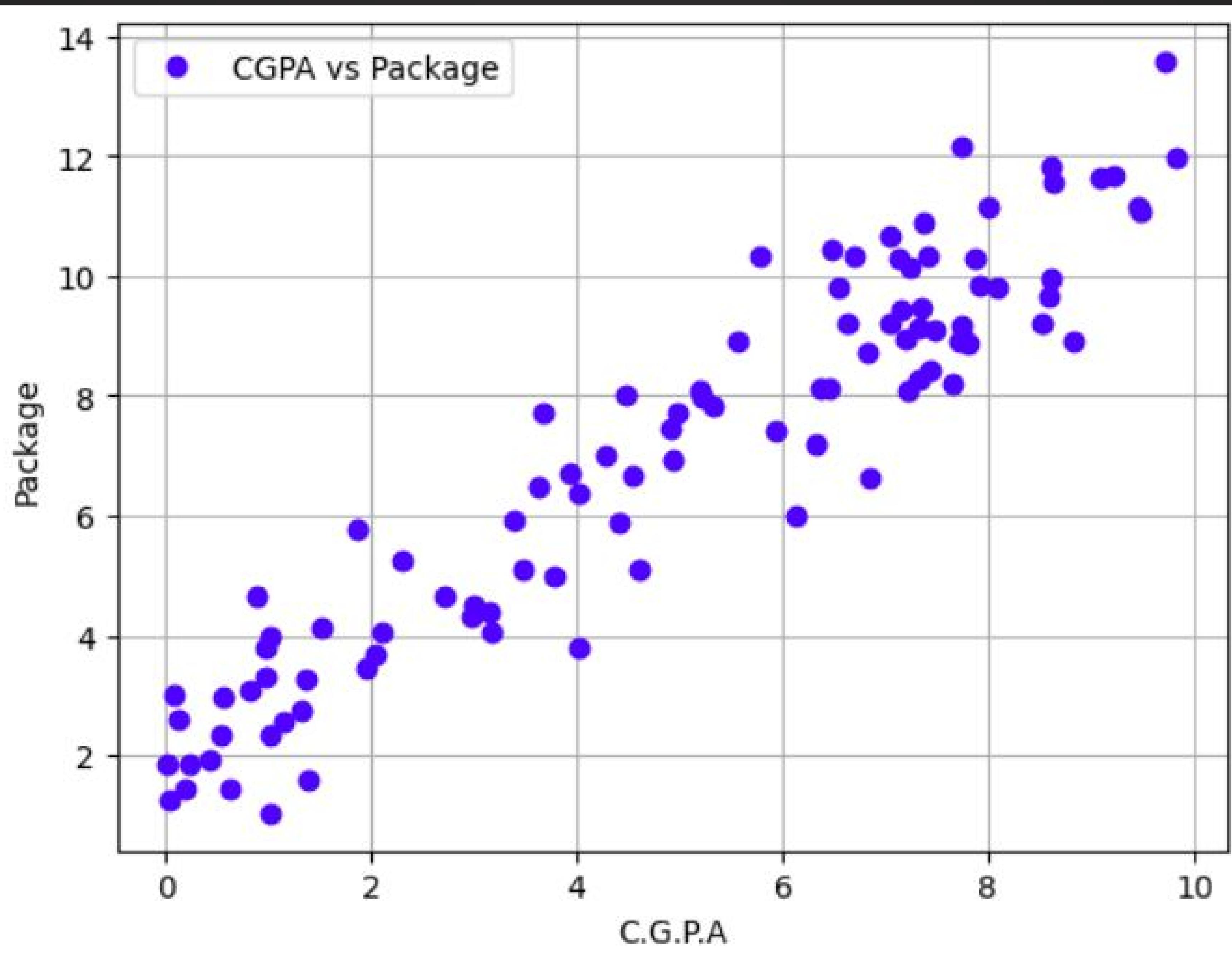
One Dimensional Data

CGPA

6.3 8.5 7.7 9.4 10

Two Dimensional Data





Three Dimensional Data

Multidimensional Data



Transaction_ID	Customer_ID	Transaction_Date	Product_ID	Quantity	Price	Payment_Method
1001	2001	2023-07-21	30001	2	19.99	Credit Card
1002	2002	2023-07-22	30002	1	9.99	PayPal
1003	2001	2023-07-23	30001	3	19.99	Credit Card
1004	2003	2023-07-24	30003	5	4.99	Debit Card
1005	2004	2023-07-25	30004	2	15.99	Credit Card

This part of statistics,
concerned with the
description and
summarization of
data, is called
descriptive statistics.

The part of statistics,
concerned with the
drawing of conclusions
from the data, is called
inferential statistics.

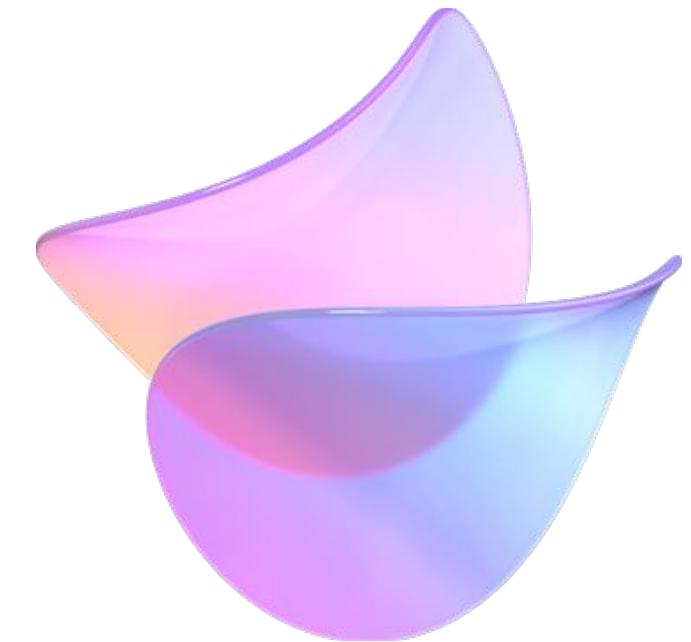
Data Analyst
AD Astra Labs

1-3 Yrs | ₹ 10-12 Lacs PA | Remote

Perform exploratory data analysis to identify trends, patterns, and relationships in the data.

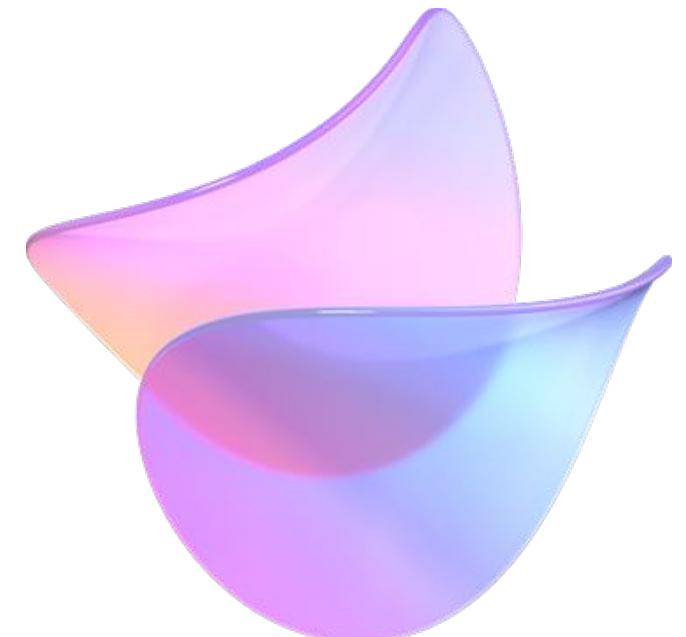
**What is a
Model?**

Huge Data

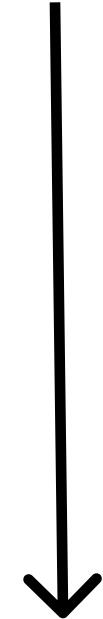


Do something

Huge Data

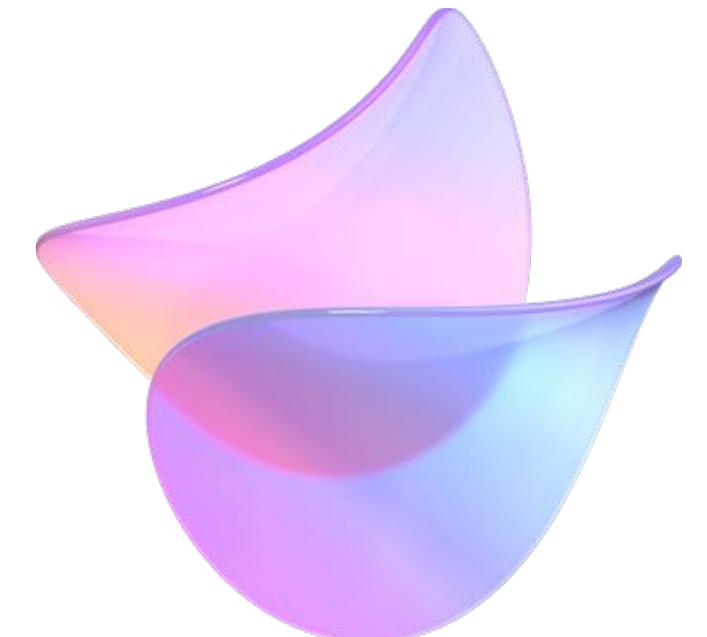


Predict Package

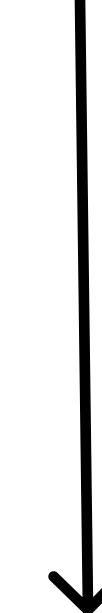


Regression

Huge Data



Job Lagegi?



Classification

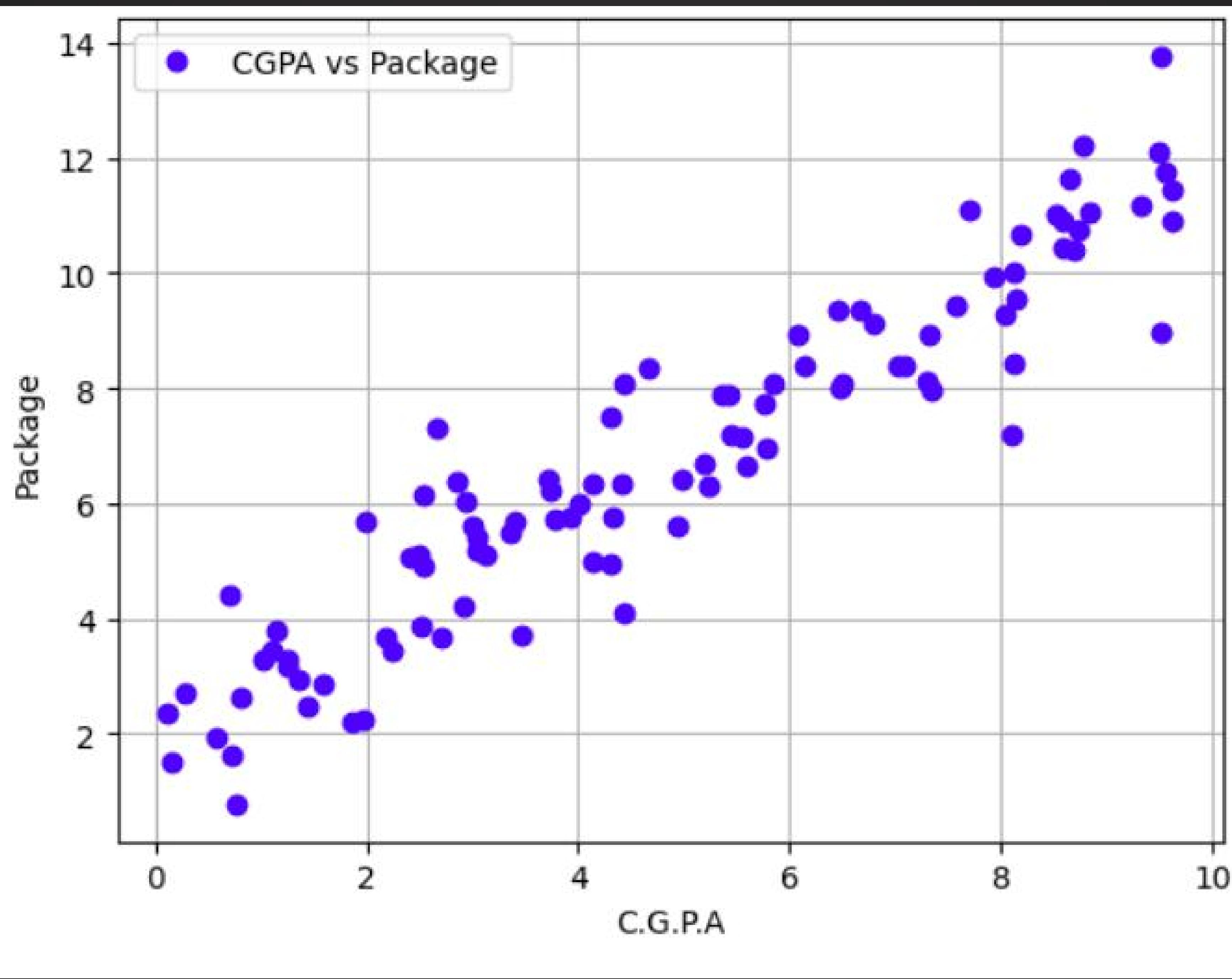
Regression

**What are
Predictions?**

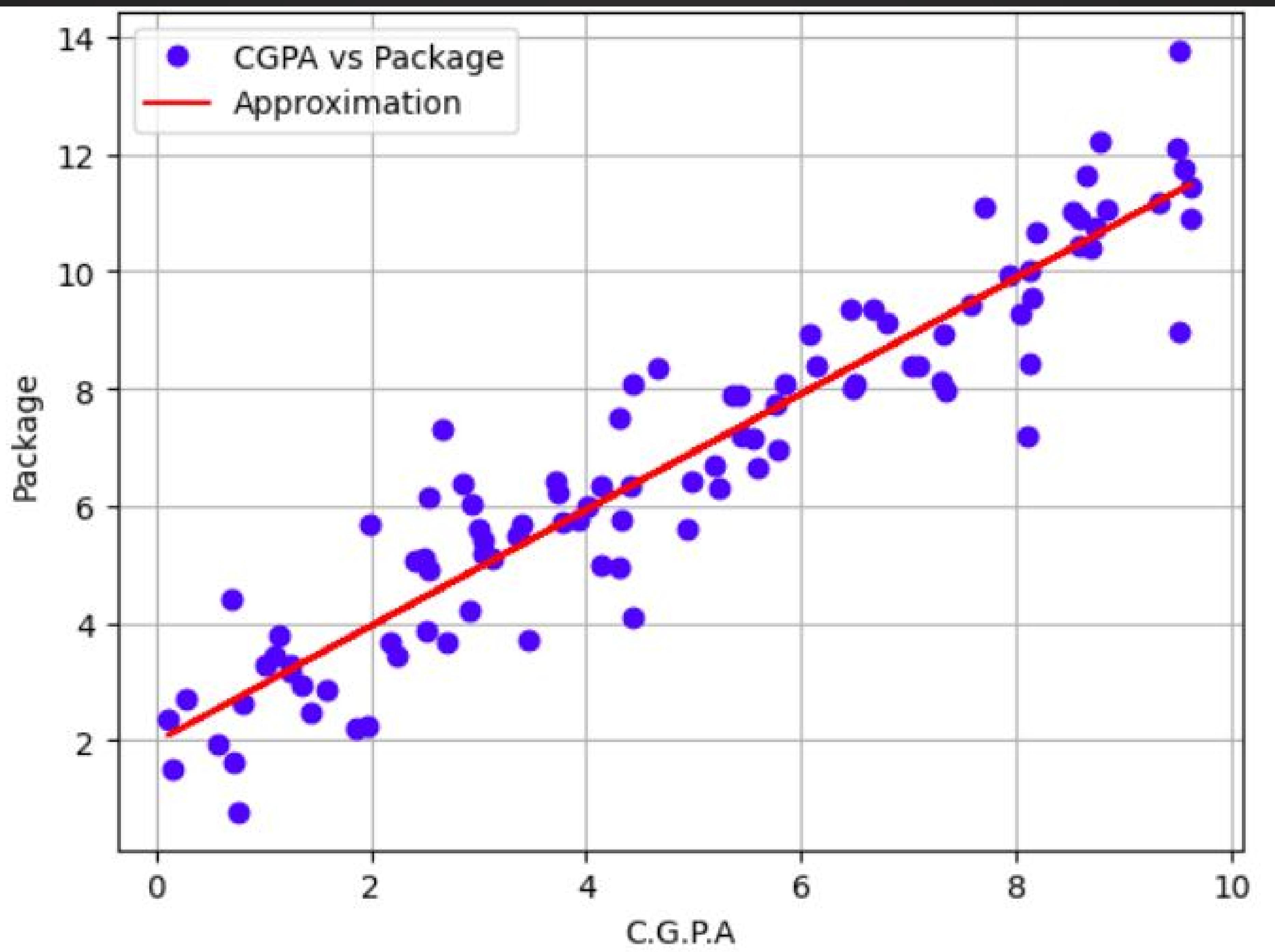
Joint Entrance Exam (JEE)

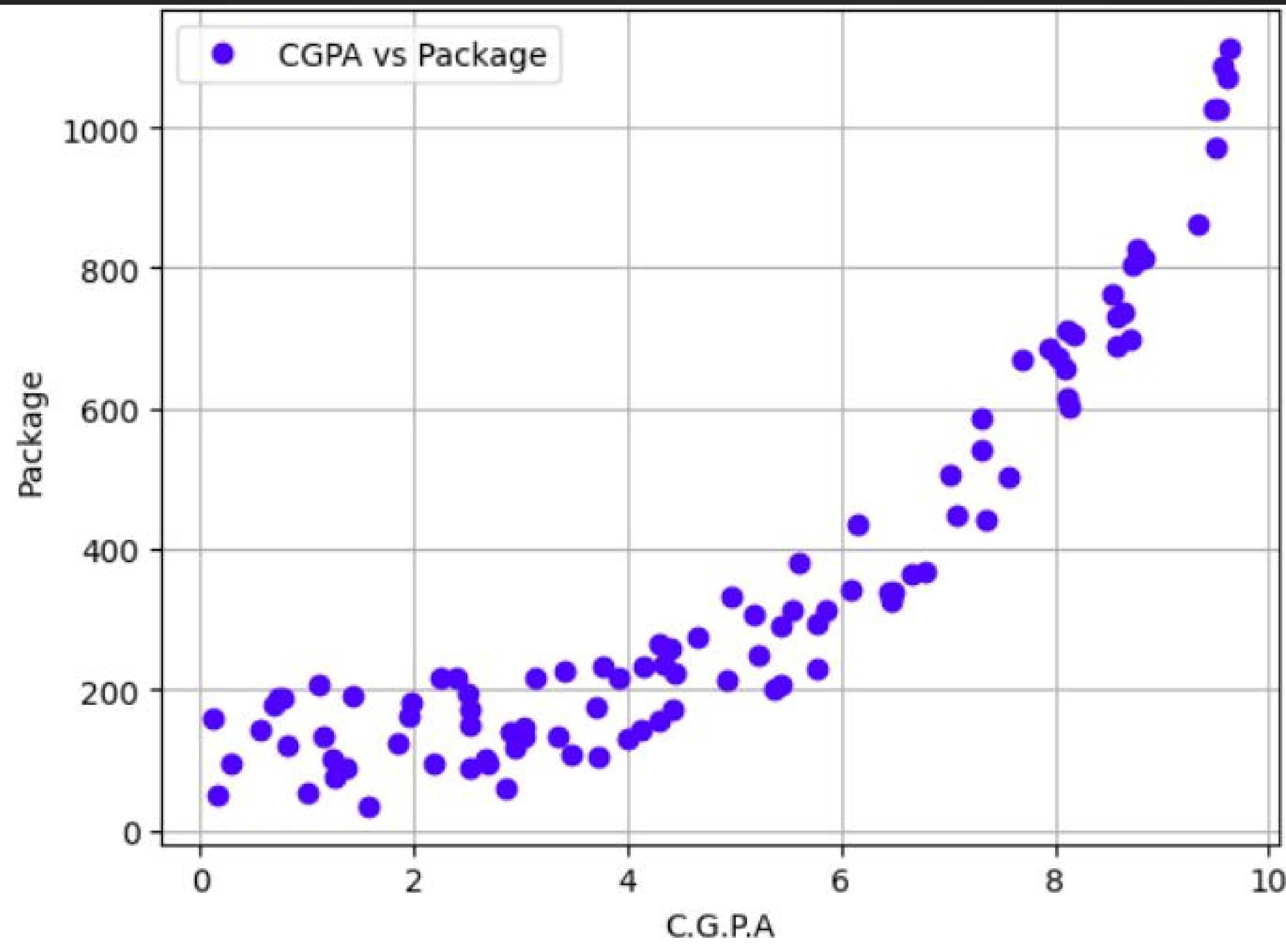


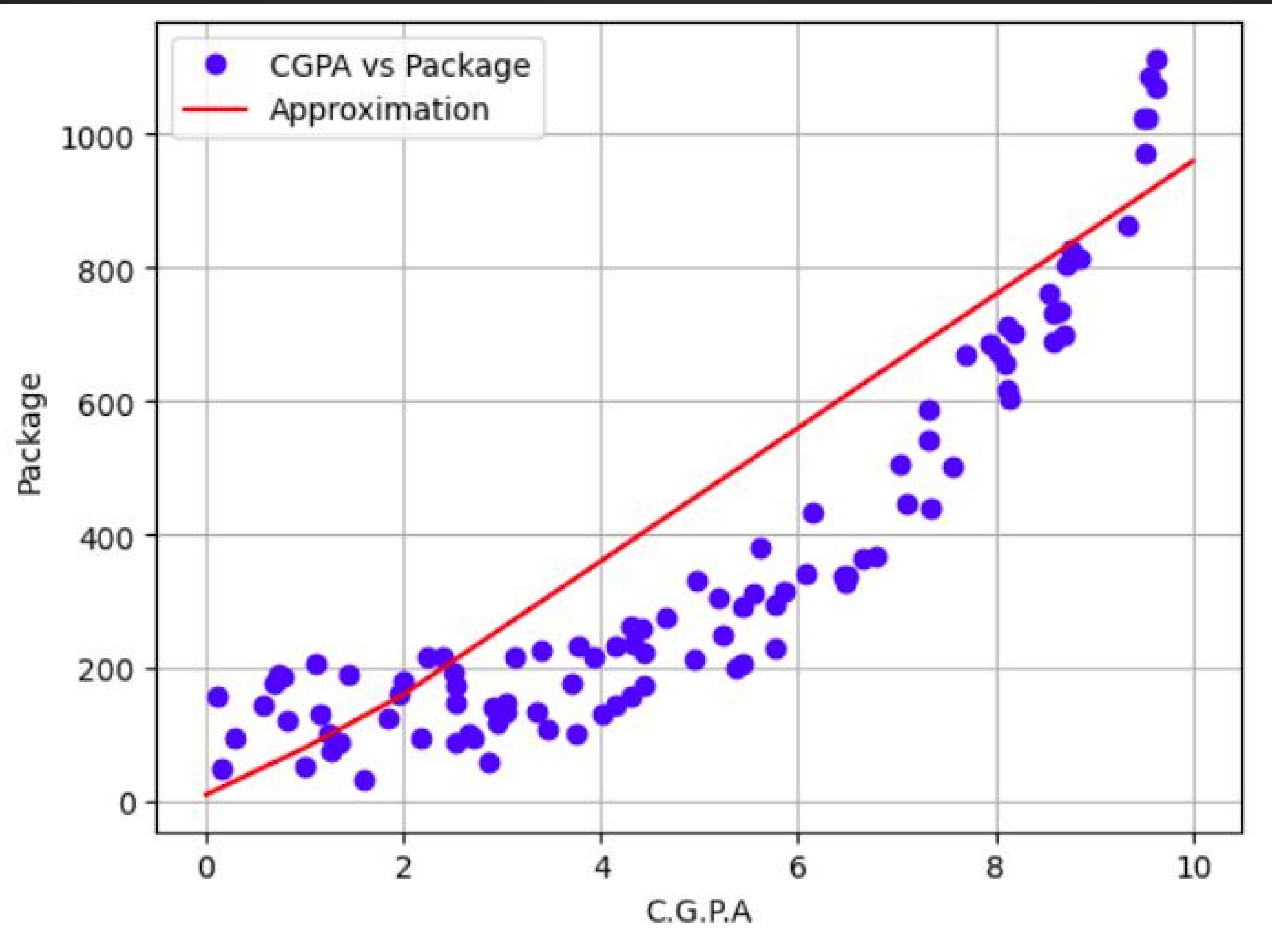
How
Predictions
are made?



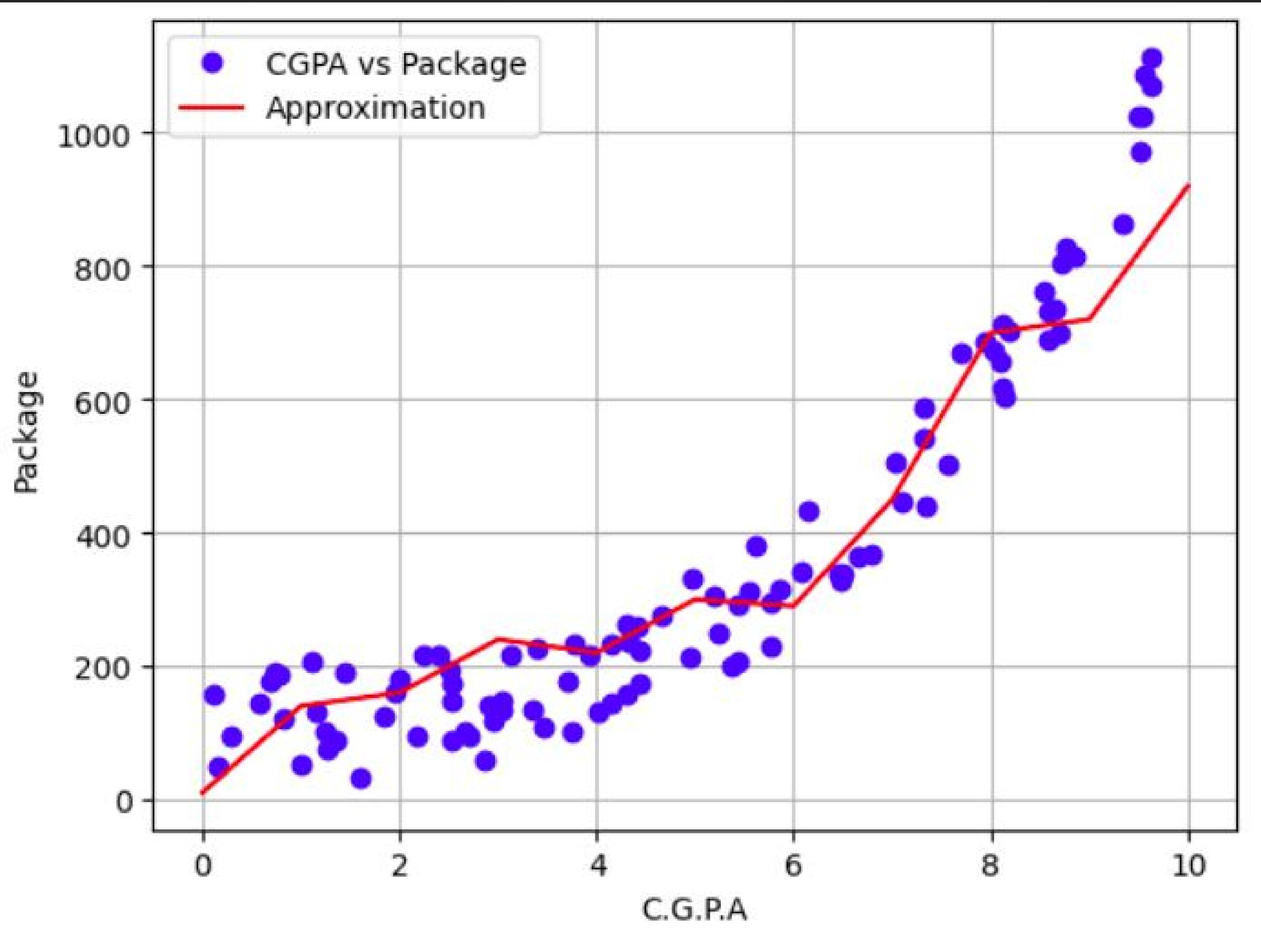
Straight
Line







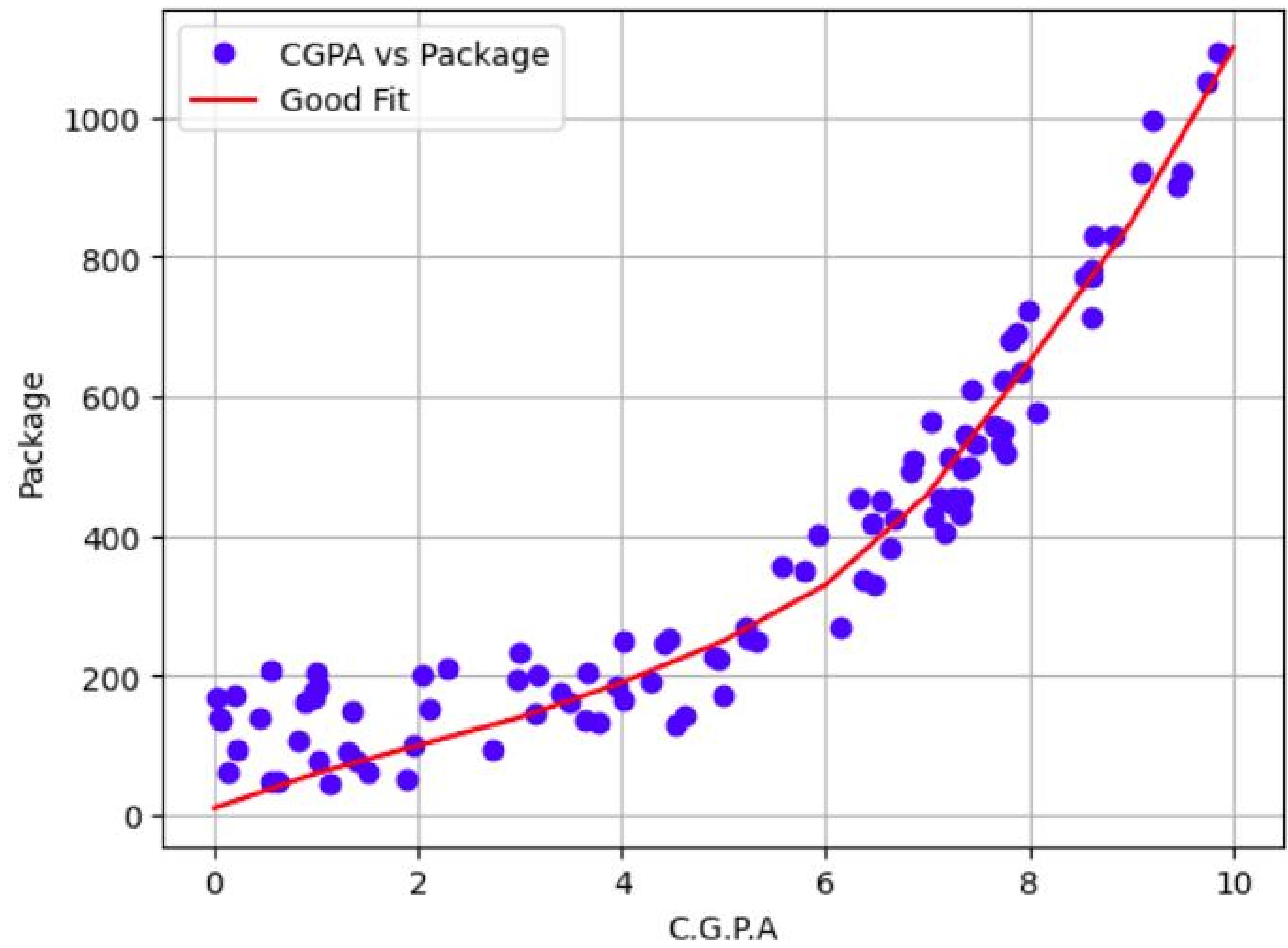
Underfitting

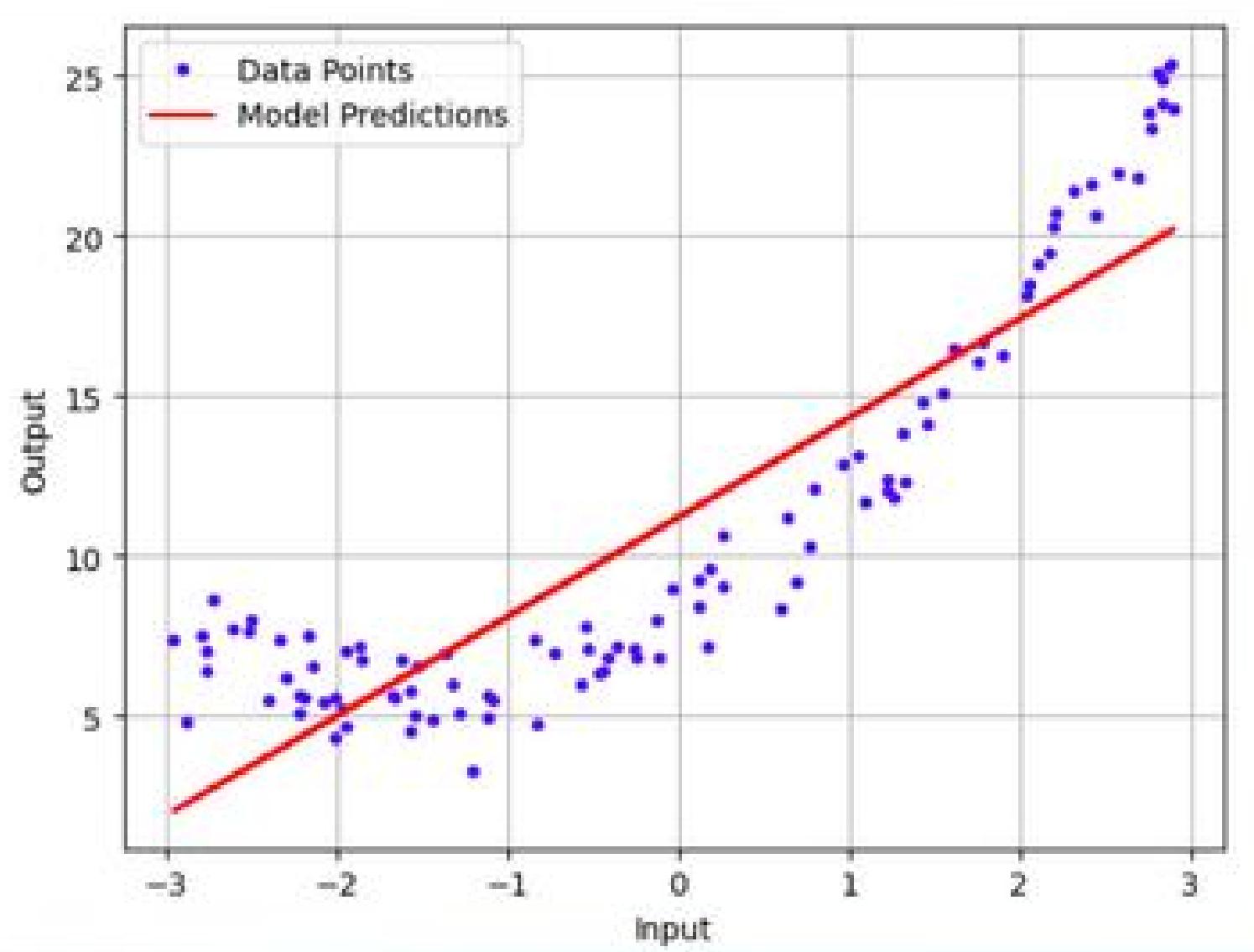


Overfitting

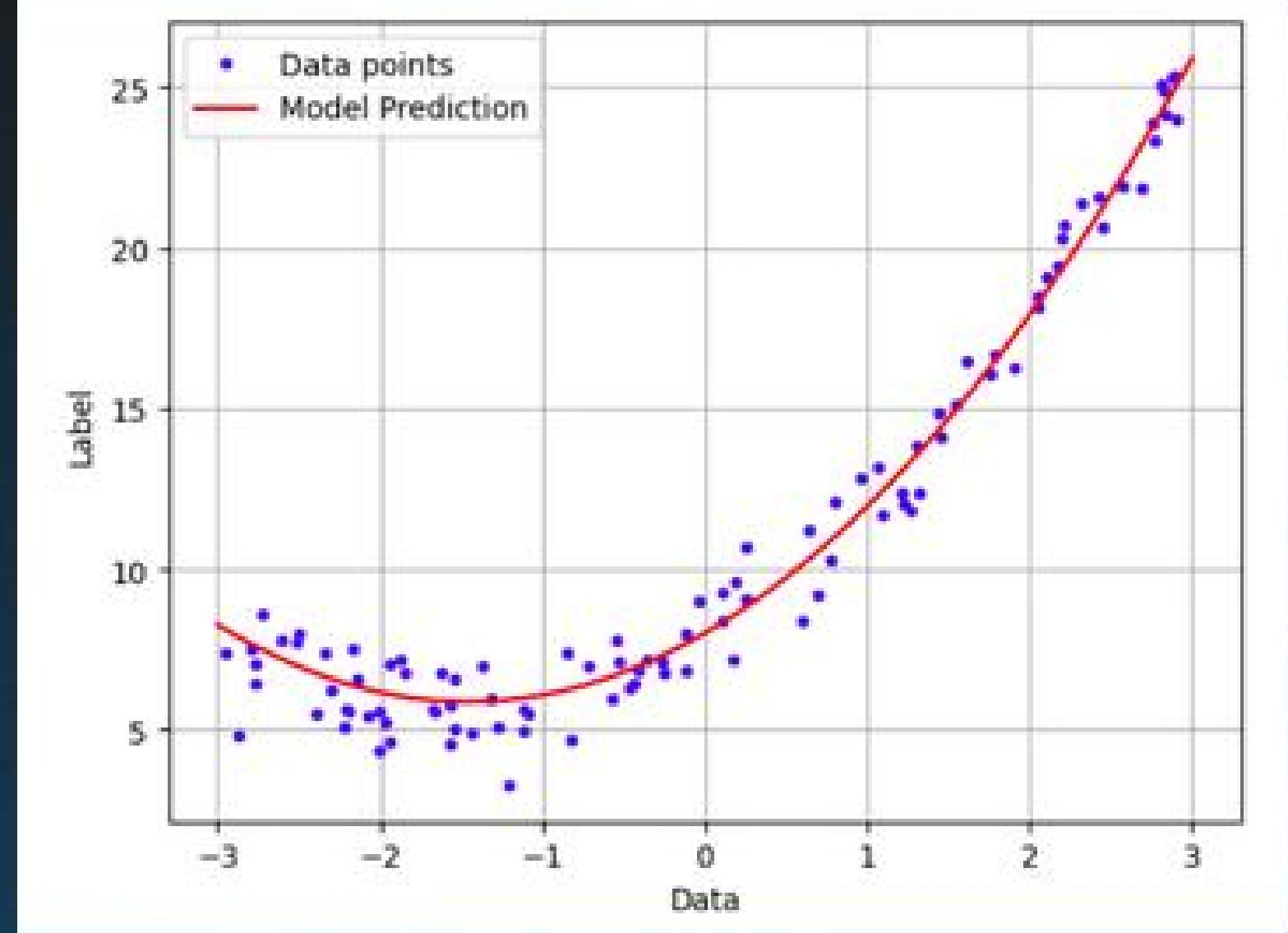
Good

Fit

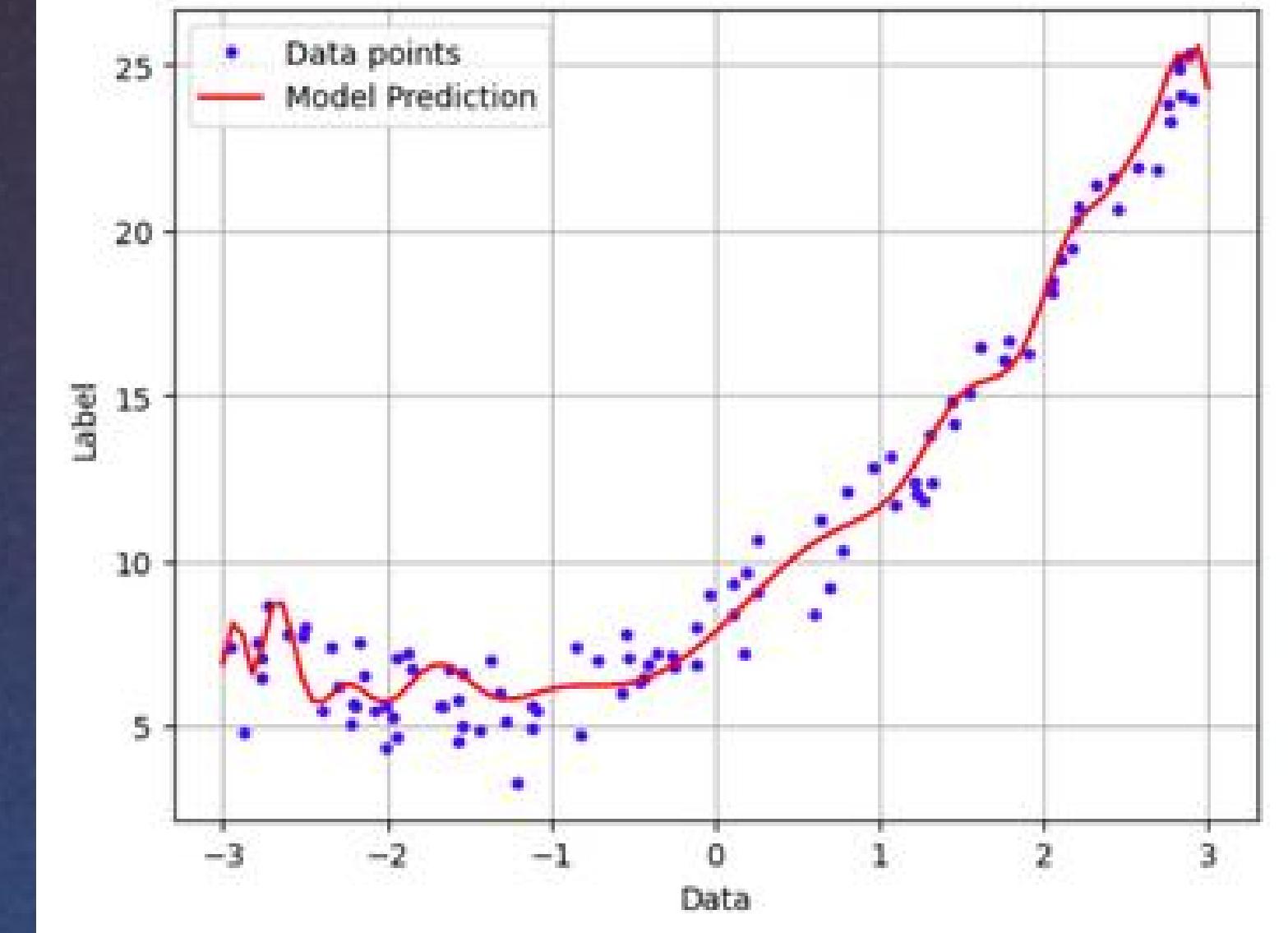




Underfitting



Good fit



Overfitting

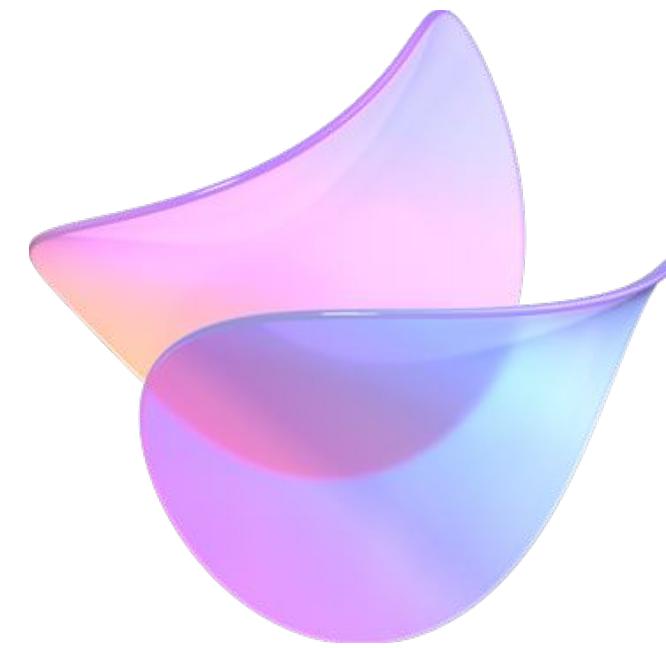
Machine Learning
Predictions are
never accurate they
are **Approximations**

ML
Systems

Supervised ML
Unsupervised ML

Supervised ML

Data



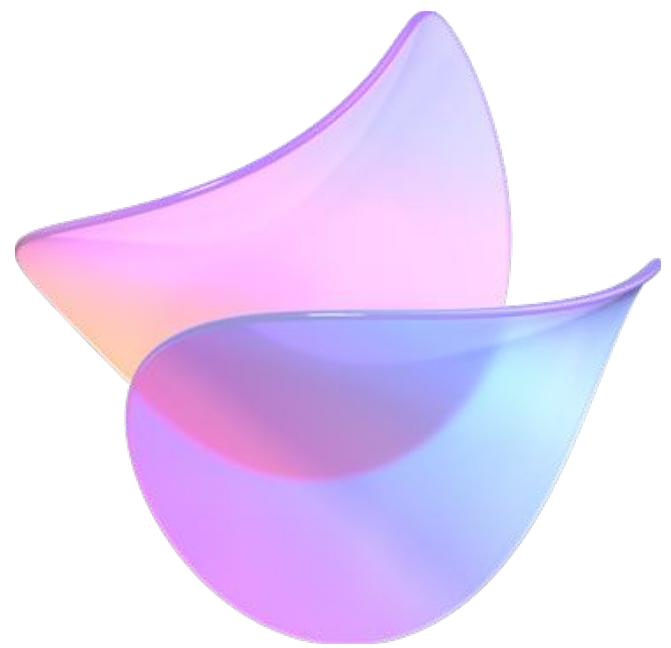
Do something



Labels

Unsupervised ML

Data



Do something



ML Project Workflow

Random Sampling

Stratified Sampling

Maths

Welcome to the
world of Machine
Learning