Machine LEARNING
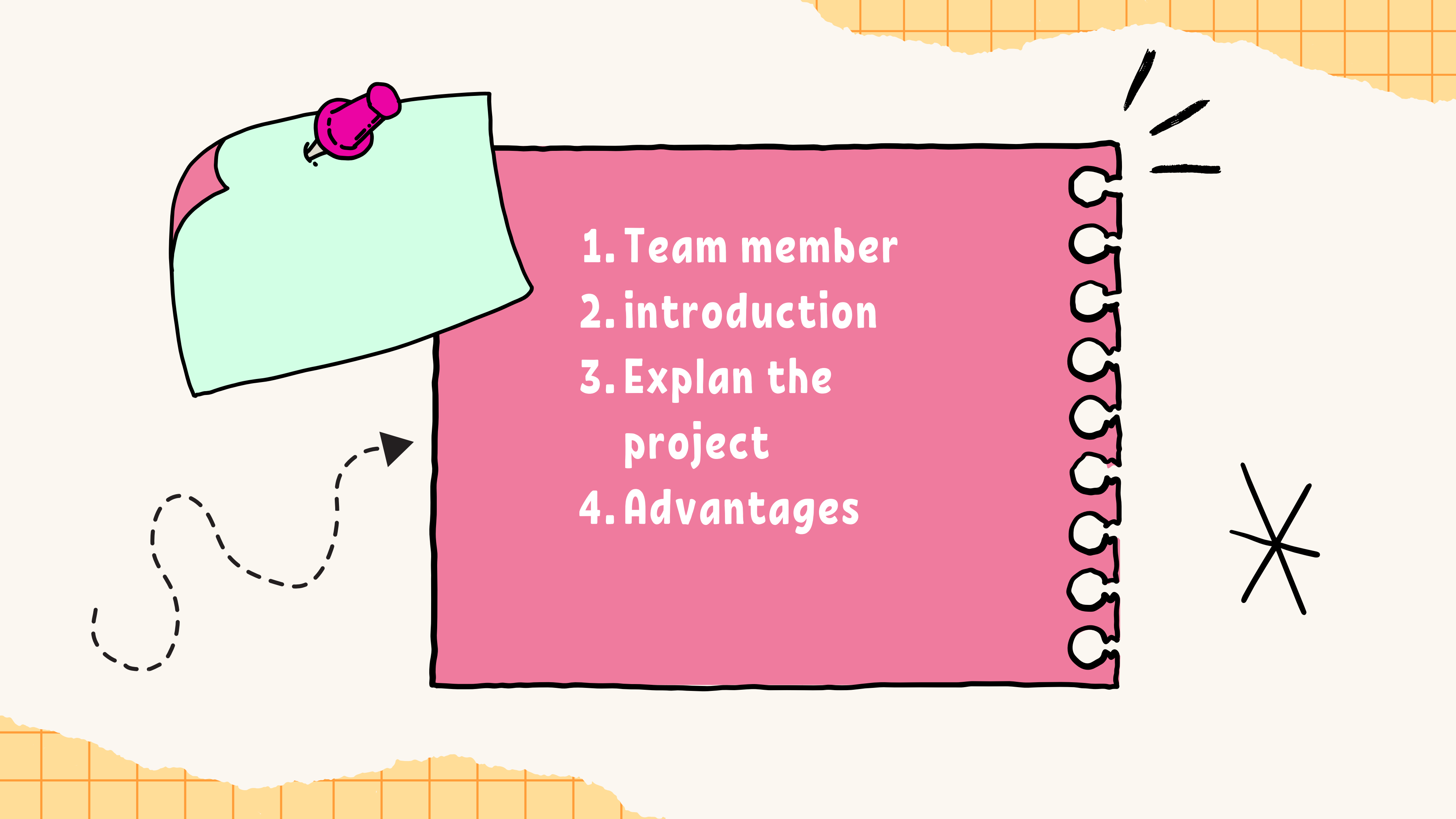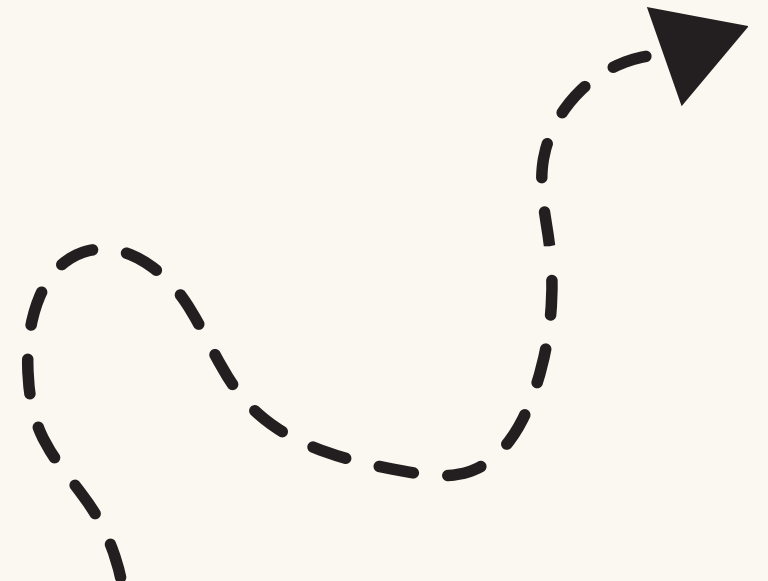
1. Team member
2. introduction
3. Explan the project
4. Advantages

Team Members

Razan mohamed kamal   2200367

Hend mofreh shafeek   2200030

yasmin Walid Gamal   2200392

Shahinaz Samy mohamed 2200369

Aya Selim mohamed   2200382

INTRODUCTION

Data mining has attracted a lot of attention in the community in recent years, being able to convert large amounts and large amounts of data into useful information and knowledge. The information and knownledge obtained can be used to apply such as market analysis, fraud detection, and customer retention, for production control and exploration science.

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from various large databases [1]. Data mining is a series of processes to explore the added value of a data set in the form of knowledge that has not been knwn manually [2]. On the off chance that the consequences of QC tests can't satisfy the acknowledgment models, the aftereffects of examination of the entire arrangement of the estimations on that day must be eliminated or should be re-dissected, and an incomplete or full re-approval of the strategy considered [12].

There are many data mining methodologies, one of which is poplar is the decision tree. Decision tree is a very interesting classification method that involves the construction of a decision tree consisting of decision nodes which are connected by branches from the root node to the leaf node (end). At the decision node the attribute will be tested, and each result will produce a branch. Each branch will be directed to another node or to the end node to produce a decision

EXPLAIN THE PROJECT

DECISION TREE IS A FLOWCHART-LIKE DIAGRAM MAPPING OUT ALL OF THE POTENTIAL SOLUTIONS TO A GIVEN PROBLEM. THEY'RE OFTEN USED BY ORGANIZATIONS TO HELP DETERMINE THE MOST OPTIMAL COURSE OF ACTION BY COMPARING ALL OF THE POSSIBLE CONSEQUENCES OF MAKING A SET OF DECISIONS.

DECISION TREES ARE MADE UP OF VARIOUS CONNECTED NODES AND BRANCHES, EXPANDING OUTWARD FROM AN INITIAL NODE. THE THREE TYPES OF NODES ARE DECISION NODES, CHANCE NODES, AND OUTCOME NODES.

- DECISION NODES ARE SQUARE-SHAPED AND REPRESENT A POINT ON THE TREE AT WHICH A DECISION CAN BE MADE. FOR EXAMPLE: SHOULD YOU HOST A BARBECUE ON SATURDAY?
- CHANCE NODES ARE CIRCLE-SHAPED AND REPRESENT A POINT ON THE TREE AT WHICH THERE ARE MULTIPLE UNCERTAIN CONSEQUENCES. FOR EXAMPLE: WHAT ARE THE CHANCES THAT IT WILL RAIN ON SATURDAY?
- OUTCOME NODES ARE TRIANGLE-SHAPED AND REPRESENT THE FINAL ENDPOINT OF A SERIES OF DECISIONS.FOR EXAMPLE: A SUDDEN DOWNPOUR RUINS YOUR BARBECUE.

Advantages of decision tree

1

It is now easy to explain, as ordinary employees can understand the subforms after a brief explanation.

2

It has value Even with little data, it is possible to form an important snapshot based on experts' descriptions of the situation (and alternative situations and costs) and their preferences for outcomes.
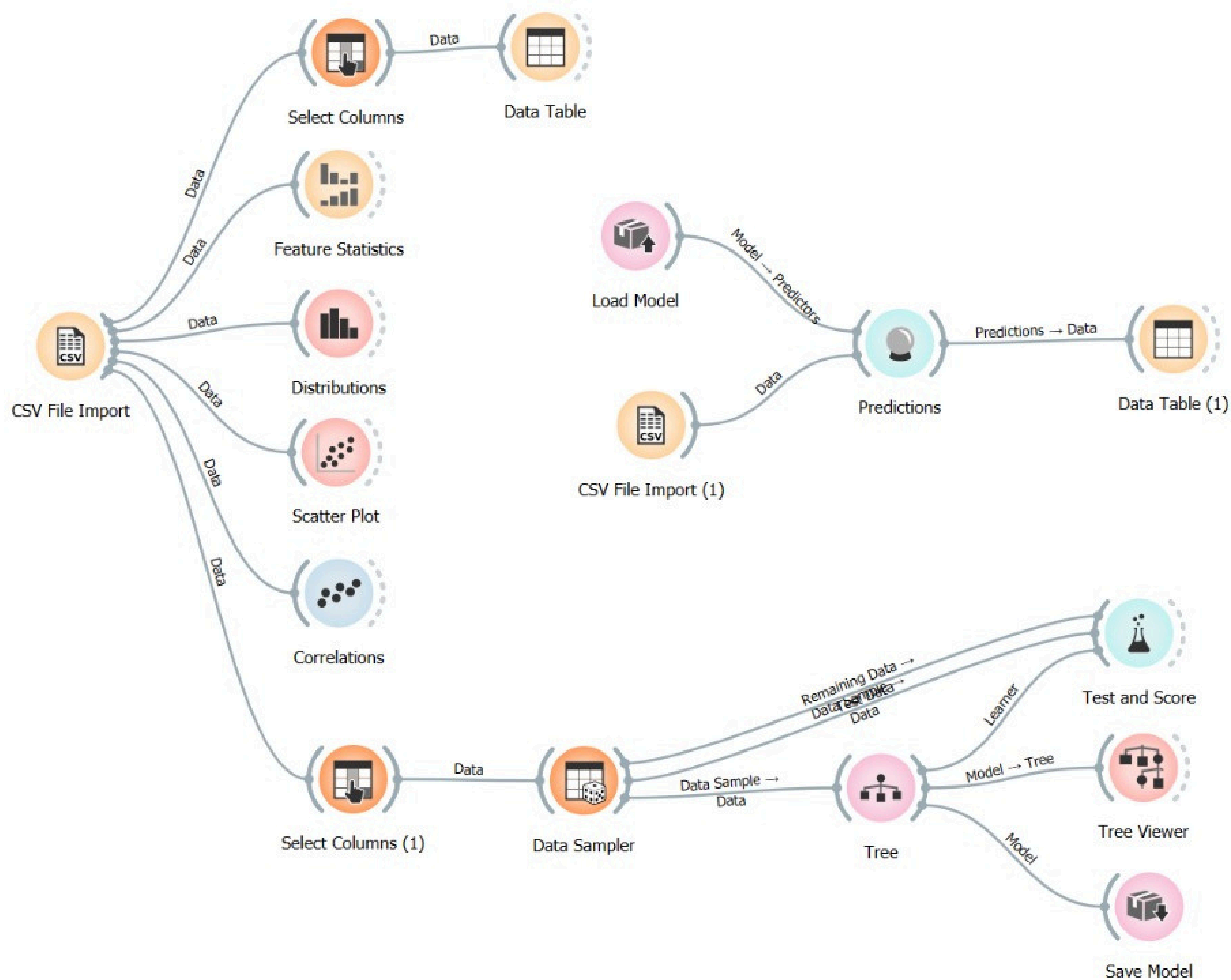
3

It also allows adding new possible scenarios

4

It contributes to determining the worst and best expected value for different scenarios

5

It can be combined with other decision techniques

6

A decision tree can be converted into decision rules

In model
By looking at what is affects the sleeping disorder it was found that
The major cause is blood pressure
A)the people with blood pressure
ranges from 128/85 to 115/75 almost all don't have sleeping disorders}
{ranges from 135/88 to 130/86 have sleeping disorders}
{ranges from 140/95 to 140/90 have sleeping disorders}
=115/75 almost every one doesn't have a sleep disorder 92%,and if they have is insomnia or sleep apnea
=120/80 almost all doesn't have a sleep disorder 96.8% ,if they have they have sleep apnea
=125/80 almost every one doesn't have a sleep disorder if have insomnia
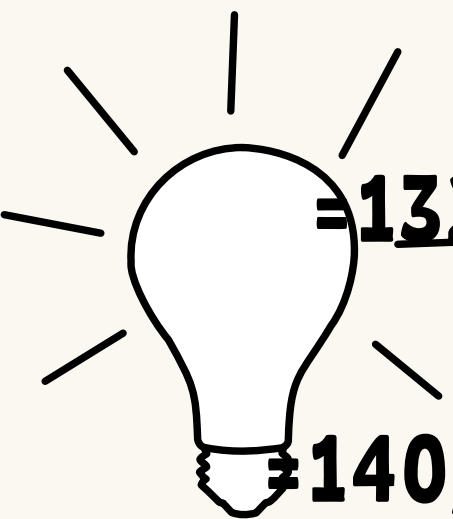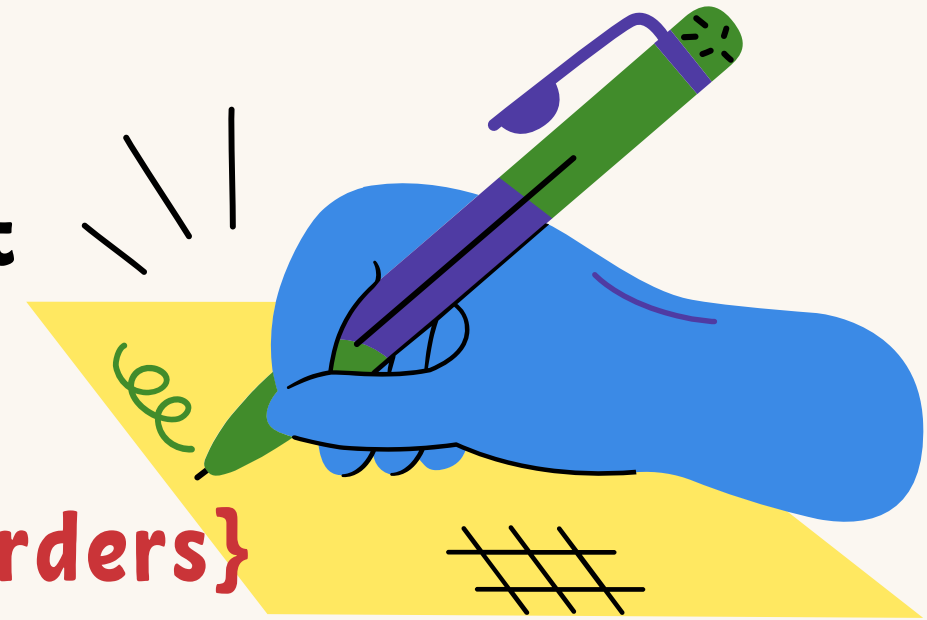===125/82 and 128/84 and 128/85 no one have a sleep disorder
=130/86 either have a insomnia or sleep apnea 50%
=132/87 either have sleep apnea or insomnia but the majority 66.7% is insomnia
=135/88 have sleep apnea only
=140/95 almost 95.7% have sleep apnea ,with very small majority having insomnia
=140/90 all have insomnia

B)1. second affecting cause is occupation with 130/85 blood pressure

Your teachers have sleep apnea ,sales person have high chances of insomnia ,all accountants have insomnia , lawyers have high chances of having no sleep disorders} text

=all accountants have insomnia

=engineers if their sleep duration is more than 7.7 hours half of them won't have sleep disorder ,But if they have less than or equal 7.7 sleep duration they have 75% chance of not having any sleep disorder

=lawyers if they have sleeping duration of more than 7.1 hours they have 93.1% chance of not having a sleep disorder ,if they sleep less than or equal to 7.7 hours they have a 75% chance of not having sleep disorder

=the age of sales person play important part if it is less than or equal to 43 they have 90% chance of having insomnia ,if they are older than 43 and sleep less than 6.3 hours all of them will have insomnia , and if they sleep more than 6.3 hours they have 66.7% chance of having insomnia
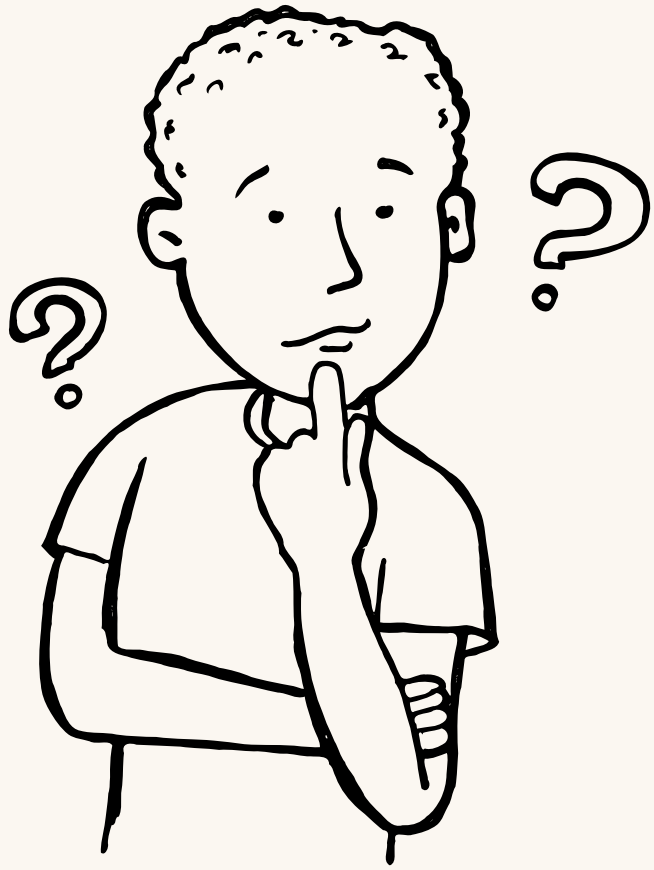
=all teachers have sleep apnea
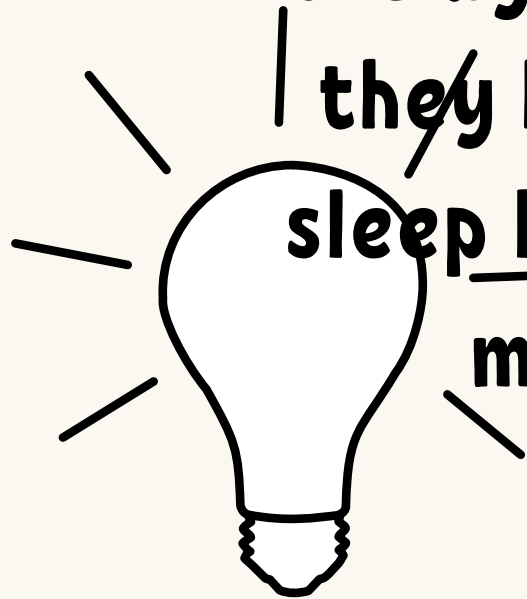
2. Age also is second leading cause with blood pressure 135/90

all have sleeping disorders}

Your paragraph=if age is less than or equal to 44 all have insomnia ,if they are older than 44 and sleep less than or equal to 6.5 hours all have insomnia ,if they sleep more than 6.5 hours they have 50 % of having insomnia
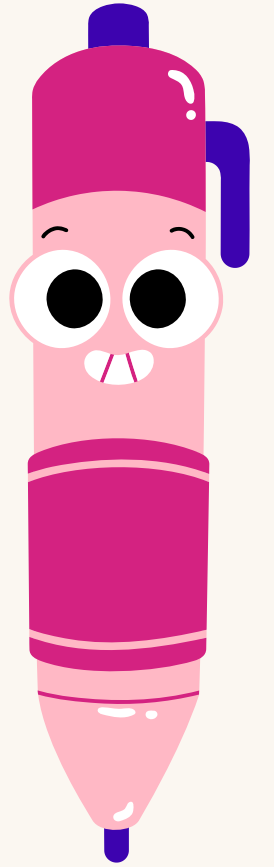
text

How To Create the Project ?

We have three parts

Part 1: Data Exploration
Part 2: Model Building
Part 3: Model Deployment

Part 1: Data Exploration

(1) Import the dataset

(2) Specify the Data Types

(3) Explore the Data:

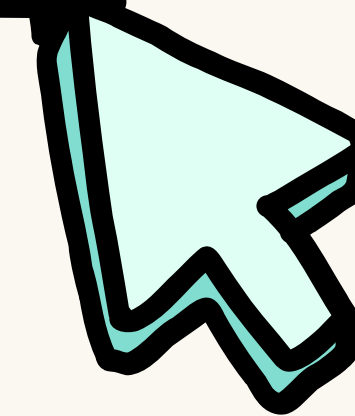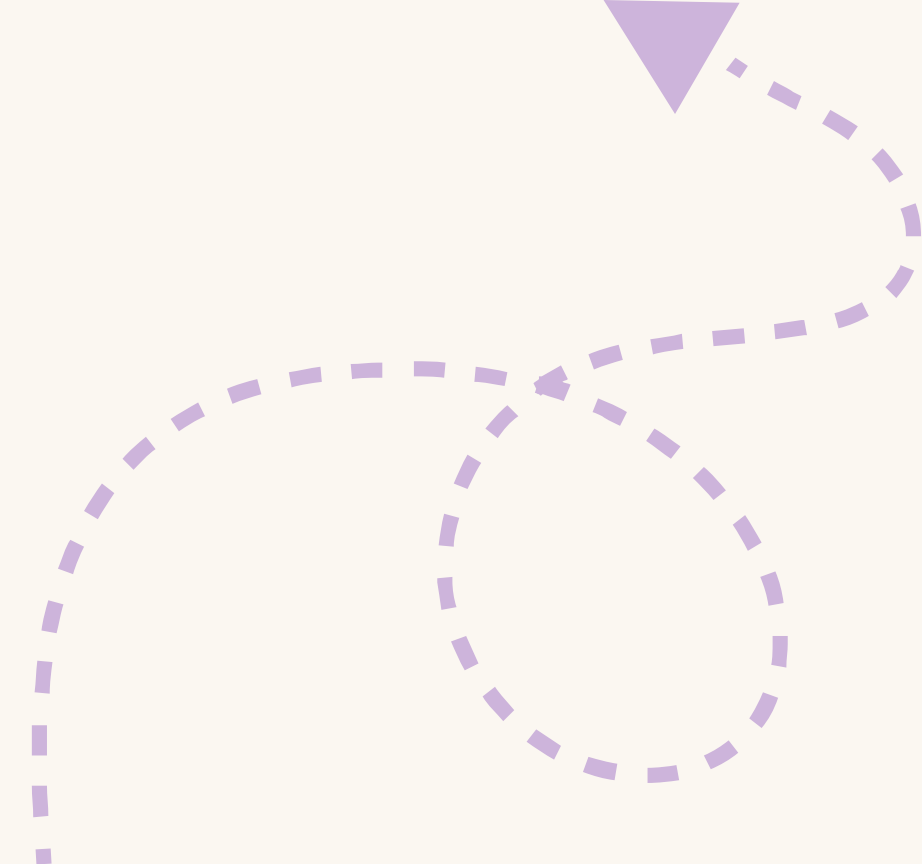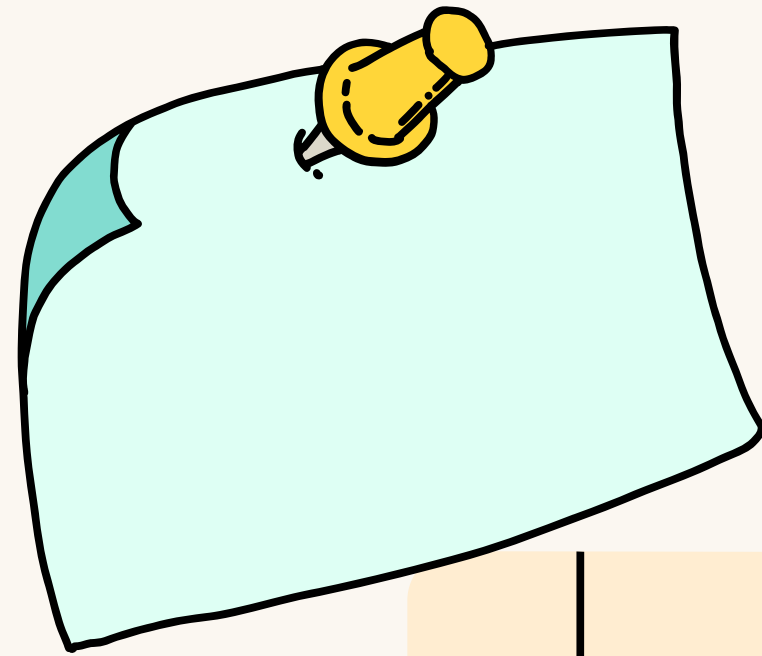a. Feature Statistics

b. Data Distributions

c. Scatter Plot

d. Correlations

Part 2: Model Building

(1) Feature Selection

(2) Split data: Train & Test

(3) Train the Decision Tree

(4) View the results

(5) Check Model Accuracy

(6) Save the Model

Part 3: Model Deployment
(1) Load the Model
(2) Load the new data
(3) View the Results
(4) Save the results

Explain the steps of the parts

1.**CSV File Import**:

1. This step involves loading data from a CSV file. The data is likely in tabular format, and this node represents the starting point of your workflow.

2. **Select Columns**:

2.After importing the data, you'll need to choose specific columns for further analysis. This could involve filtering out unnecessary information or focusing on relevant features.

3.**Feature Statistics**:

3.Here, you'll compute statistics related to the features in your dataset. Common statistics include mean, standard deviation, and correlation coefficients. These insights help you understand the data better.

4. **Data Table**:

This node likely represents the processed data after selecting columns and computing feature statistics. It's the intermediate step before model building.

5. **Load Model**:

In the machine learning context, this step involves loading a pre-trained model. It could be a regression model, classification model, or any other type of ML model.

6. **Predictions**:

Once the model is loaded, you can make predictions on new data. For example, if it's a classification model, you'd predict class labels based on input features.

7. **Data Sampler**:

This path seems to diverge from the main flow. It suggests that you're sampling data (perhaps for training purposes). Sampling helps create a representative subset of your dataset.

8. **Tree**:

Your paragraphThe "Tree" node likely refers to a decision tree model. Decision trees are used for classification and regression tasks. They split data based on feature values to make predictions. text

9. **Test and Score**:

After building the tree model, you'll evaluate its performance using test data. Metrics like accuracy, precision, recall, or F1-score are commonly used.

10. **Save Model**:

If the tree model performs well, you'll save it for future use. Saving models allows you to reuse them without retraining.

THANK YOU!