

IBM NALAIYA THIRAN

ASSIGNMENT 1

Name: RANJITH K S

Roll No: 737819CSR153

Challenge:

Upload the dataset to Cognos Analytics, explore and visualize the data.

Dataset: [50-Startup-Dataset](#)

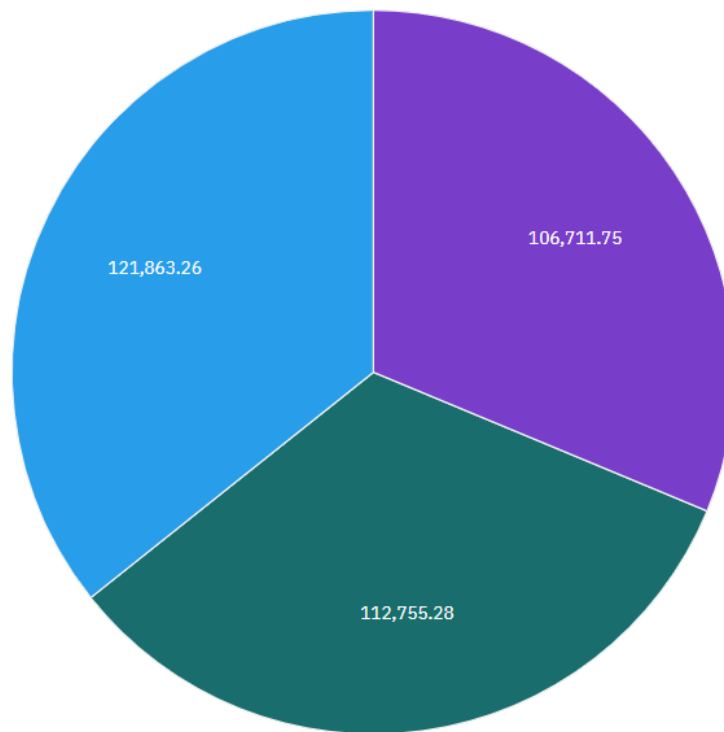
Exploration and Visualization:

1. Profit By State

Profit by State 4

State

California New York Florida

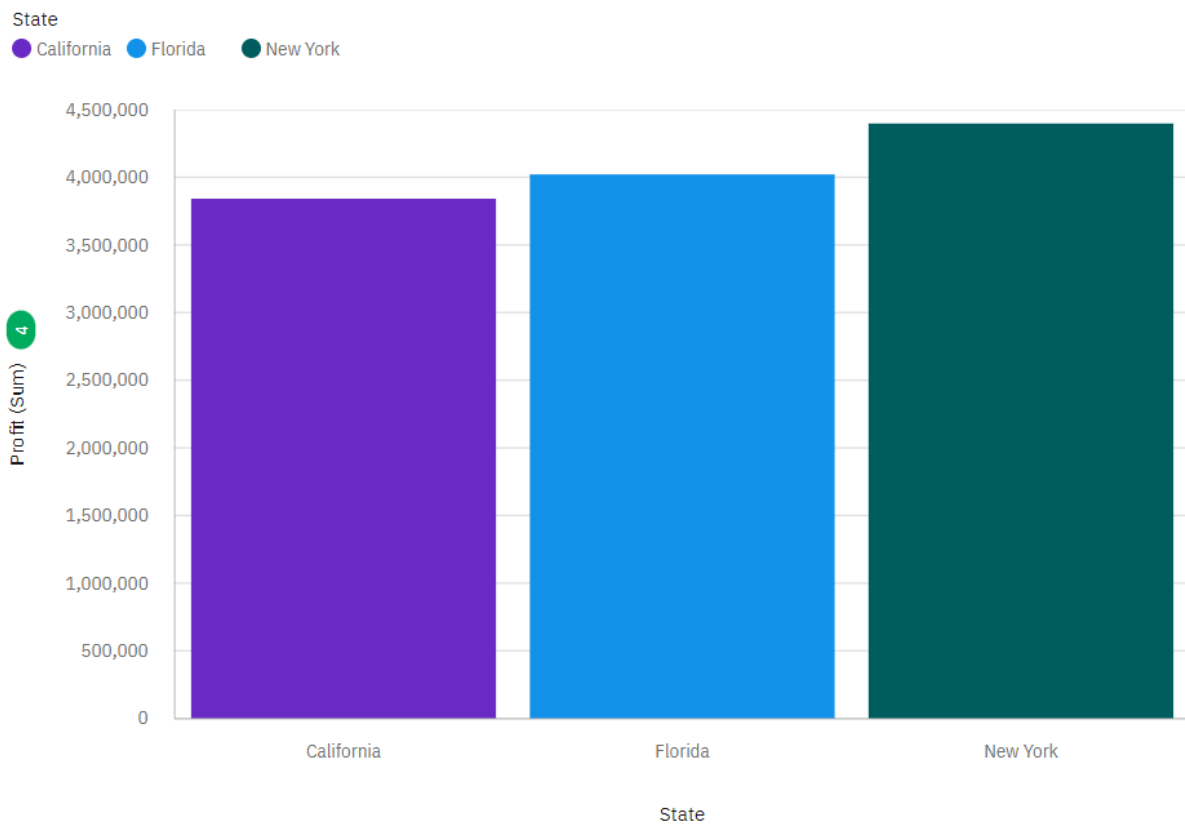


Inference:

- Over all **states**, the average of **Profit** is nearly 114 thousand.
- On an average each state contributes over 100 thousand.

2. Profit by State Colored by State

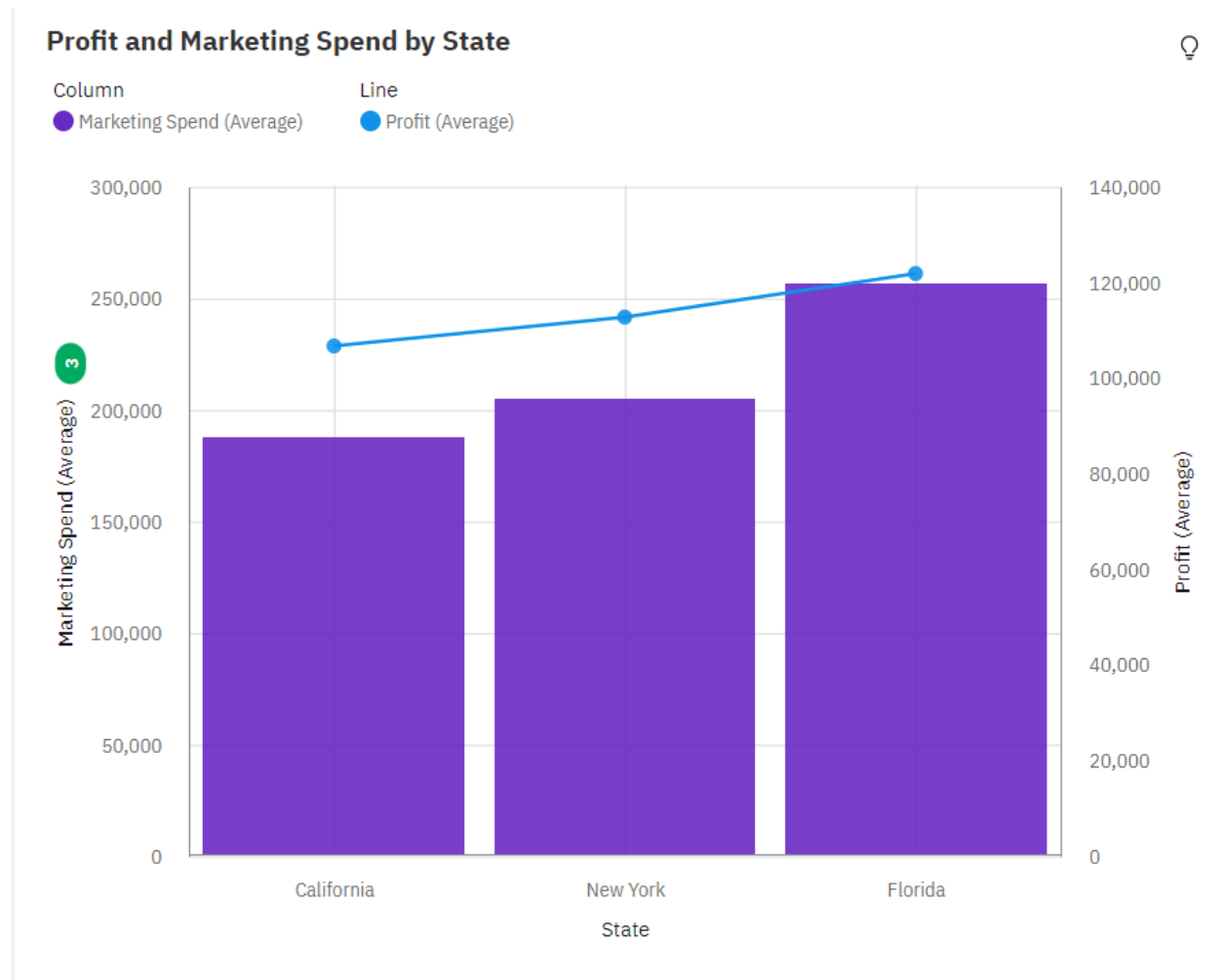
Profit by State colored by State



Inference:

- Across all **states** and **states**, the sum of **Profit** is over 12 million.
- For **Profit**, the most significant value of **State** is New York, whose respective **Profit** values add up to almost 4.4 million, or 35.9 % of the total.

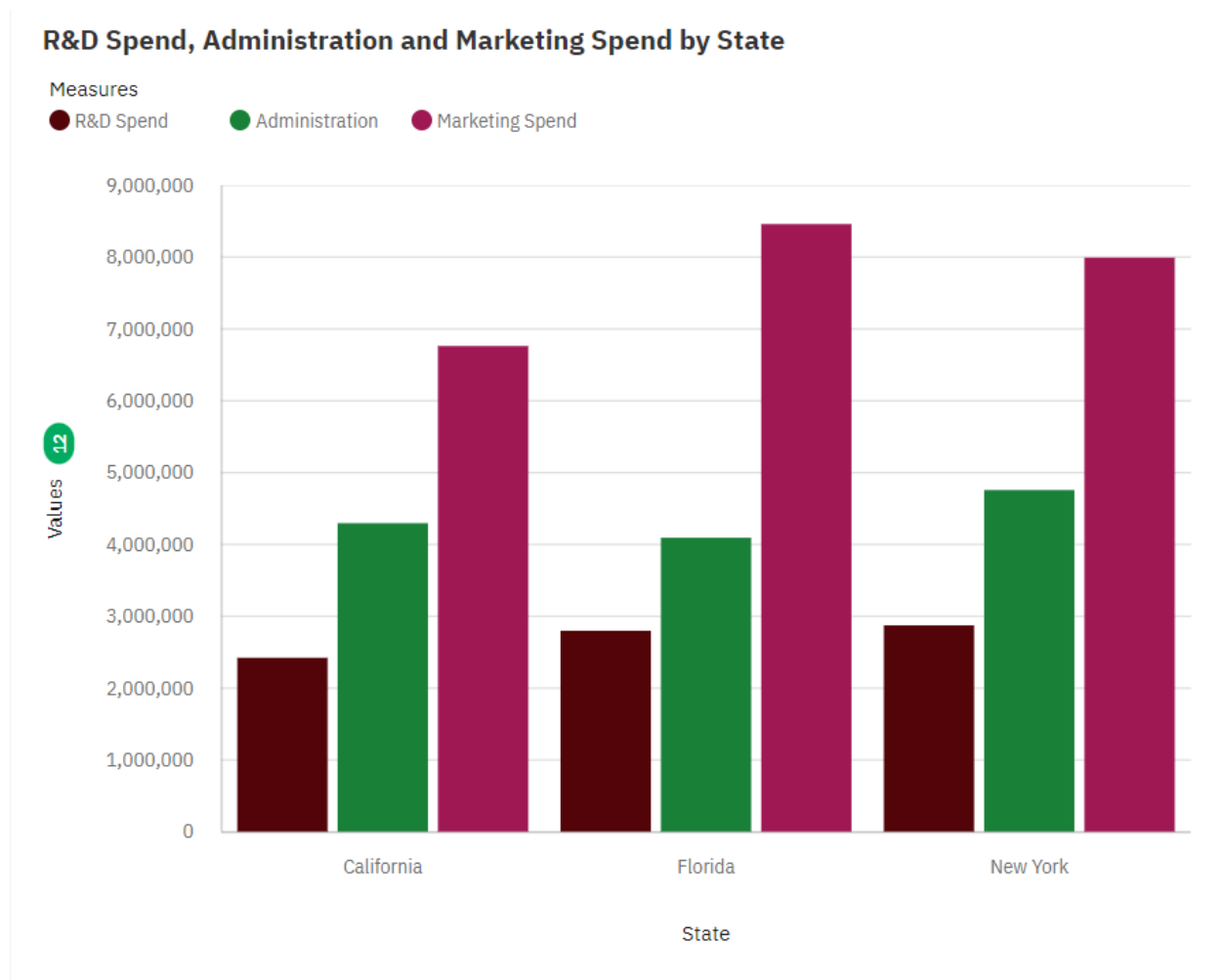
3. Profit and Marketing Spend by State



Inference:

- Over all **states**, the average of **Marketing Spend** is almost 215 thousand.
- From the visualization, we can see that marketing spend and profit is correlated.

4. R&D Spend, Administration and Marketing Spend by State



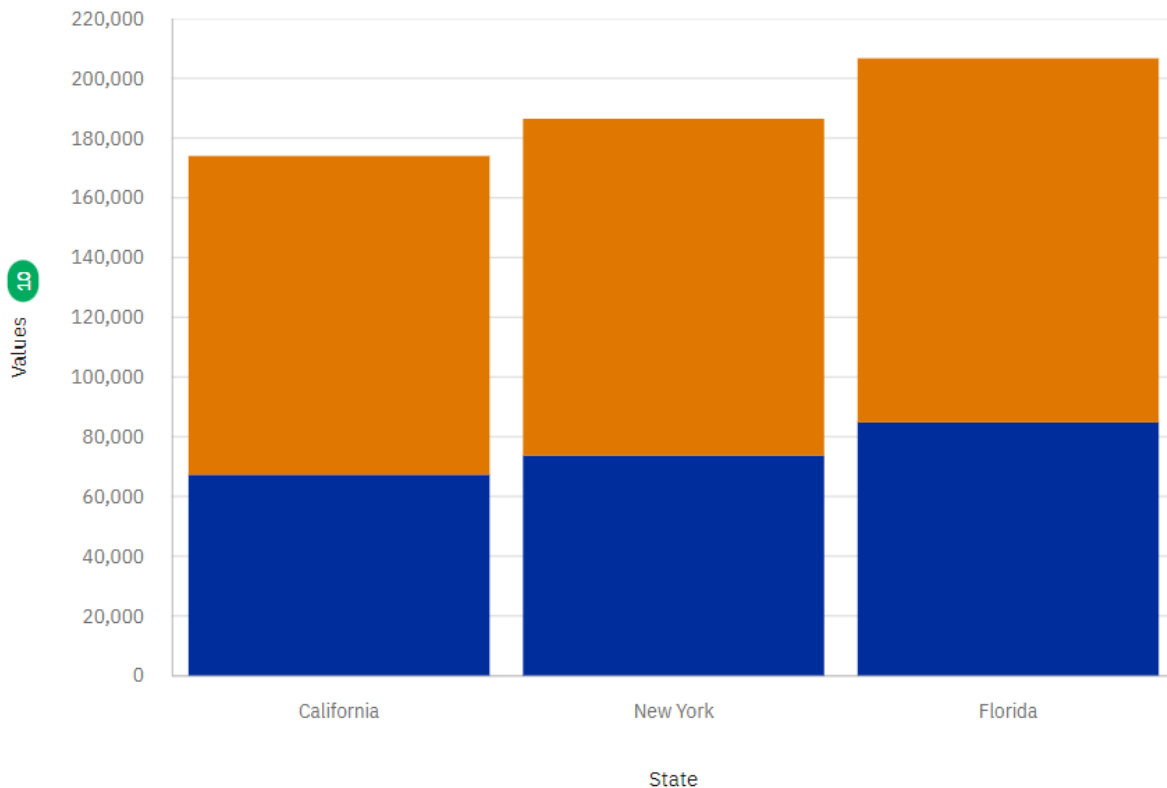
Inference:

- **R&D Spend** ranges from over 2.4 million, in California, to almost 2.9 million, in New York.
- **Administration** ranges from almost 4.1 million, in Florida, to nearly 4.8 million, in New York.
- **Marketing Spend** ranges from nearly 6.8 million, in California, to nearly 8.5 million, in Florida.

5. R&D Spend and Profit by State

R&D Spend and Profit by State

Measures
● R&D Spend ● Profit



Inference:

- The average values of **R&D Spend** range from over 67 thousand, occurring in California, to almost 85 thousand, in Florida.
- The average values of **Profit** range from almost 107 thousand, occurring in California, to nearly 122 thousand, in Florida.
- R&D Spend is correlated with Profit by 90%.

6. Profit Prediction Using Decision Tree:

Profit



Tree sunburst

Tree diagram

Rules

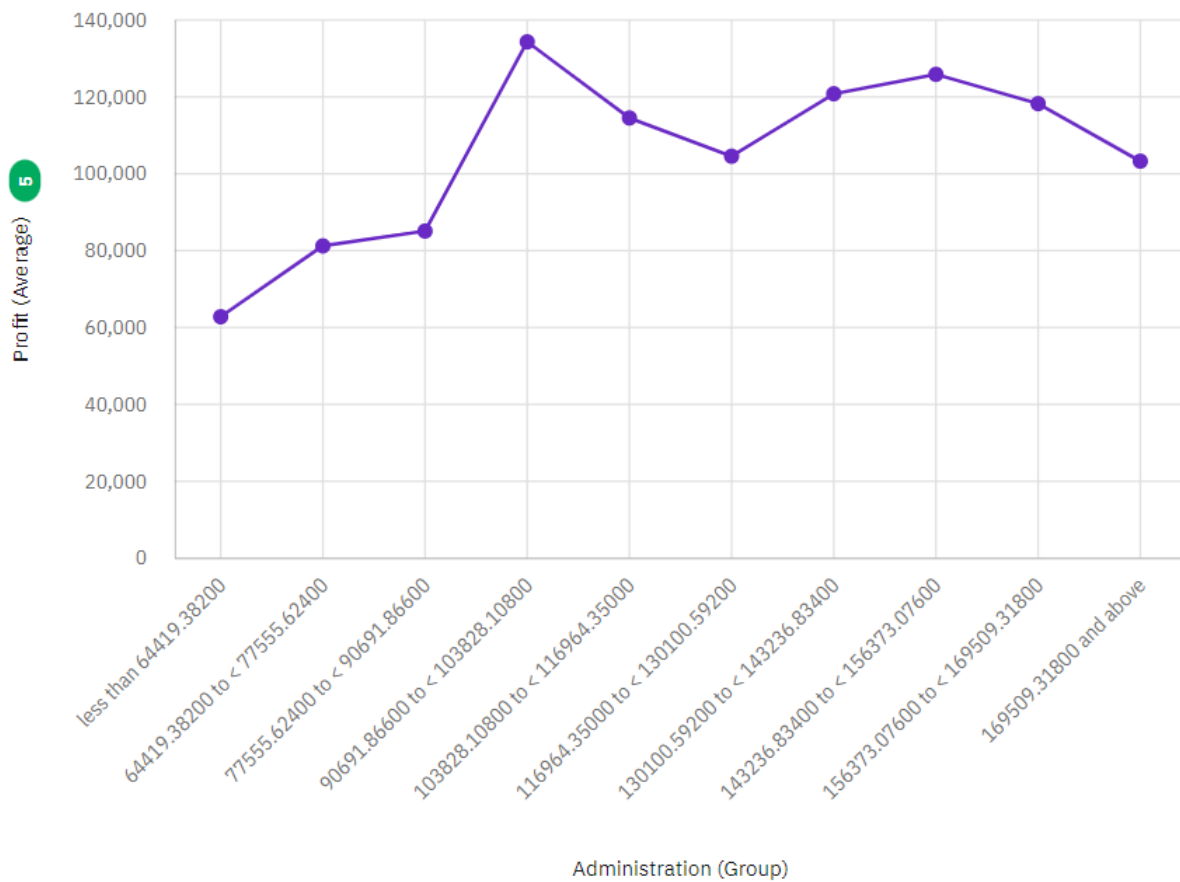
Predicted value	↓	Rules	Records
149,440.00		R&D Spend \geq 86,420	45 (42%)
87,869.50		R&D Spend $<$ 86,420	63 (58%)

Inference:

- **R&D Spend** predicts **Profit** with a strength of 61.2%.

7. Profit by Administration (Group)

Profit by Administration (Group)



Inference:

- **Administration (Group)** weakly affects **Profit** (11%).

8. Data Relationship Diagram:

Explore data relationships

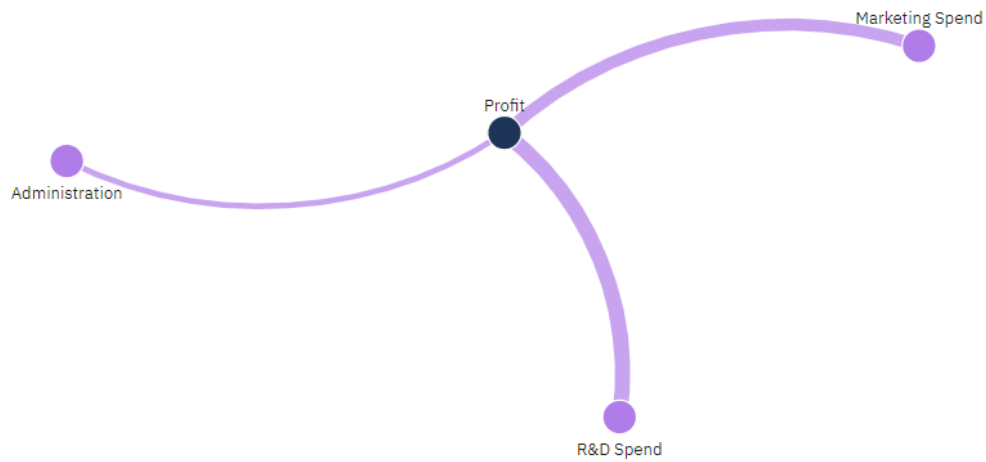
50_Startups.csv

[Reset to original](#)

Q Profit

×

Edit diagram ▾



Select single or multiple nodes to see visualizations.

Relationship diagram ⓘ

10% —●————●— 100%

Inference:

- Profit is related highly(strength: 75%) with R&D Spend.
- Profit is related moderately(strength: 56%) with Marketing Spend.
- Profit is hardly related with Administration (strength: 16%).

Conclusion:

The dataset was added to Cognos Analytics and successfully explored and visualized the data.