

A NEW HINT TO TRANSPORTATION - ANALYSIS OF THE NYC BIKE SHARE SYSTEM

TASK:

Upload the dataset to Cognos Analytics, Explore and Visualize the 50-startups dataset.

DATASET USED:

This particular dataset holds data from 50 startups in New York, California, and Florida.

The

features in this dataset are **R&D spending**, **Administration Spending**, **Marketing Spending**, and **Location features**, while the target variable is: **Profit**.

- 1. R&D spending:** The amount which startups are spending on Research and development.
- 2. Administration spending:** The amount which startups are spending on the Admin panel.
- 3. Marketing spending:** The amount which startups are spending on marketing strategies.
- 4. State:** To which state that particular startup belongs.
- 5. Profit:** How much profit that particular startup is making.

SOLUTION:

The screenshot shows the IBM Cognos Analytics with Watson interface. On the left, there's a 'Hello. Welcome to Watson.' message and a 'Watch video' button. Below it is a 'Quick launch' section with a downward arrow icon. In the center, there's a 'Documents' section with a tree view showing 'Assignment-1' and '50_Startups.csv'. A preview of the CSV file is displayed, showing columns for R&D Spend, Administration Spend, Marketing Spend, and Profit. To the right, there's a search bar and a sidebar with 'Favourites' (Recents, Applications, Desktop), 'Tags' (Red, Orange, Yellow, Green, Blue, Purple, Gray, All Tags...), and an 'Information' section for the selected CSV file. The bottom right corner has a 'Present data' button.

The screenshot shows the IBM Cognos Analytics with Watson interface. At the top, there's a navigation bar with 'IBM Cognos Analytics with Watson', a 'Content' dropdown, a user icon with '359', a search bar, and various status icons. Below the navigation is a header with 'Content' and a 'New' button. Underneath is a sub-header with 'My content', 'Team content', and 'Samples'. The main area displays a grid of datasets. One dataset, '50_Startups.csv', is highlighted with a red box. Other datasets shown include 'IPL Ball-by-Ball 2008-2020.csv', 'IPL Matches 2008-2020.csv', 'Olympic Events.xlsx', 'Olympic Medals.xlsx', and 'sales.csv'. Each dataset card includes a file name, last accessed date, and download links.

As highlighted by the red box, the dataset *50_Startups.csv* was successfully **uploaded** onto the IBM Cognos platform.

EXPLORATION AND VISUALIZATION OF THE DATASET:

As a standard step for exploration and visualization, A new exploration was created so that the dataset can be analyzed and visualized to provide further insights.

The screenshot shows the 'New' menu in the IBM Cognos Analytics with Watson interface. The 'Exploration' option is selected and highlighted in blue. Other options in the menu include 'Data module', 'Dashboard', 'Report', and 'Story'. The left sidebar shows navigation links like 'Home', 'New', 'Upload data', 'Content', 'Recent', and 'Manage'.

The data source was then added to carry out the exploration.

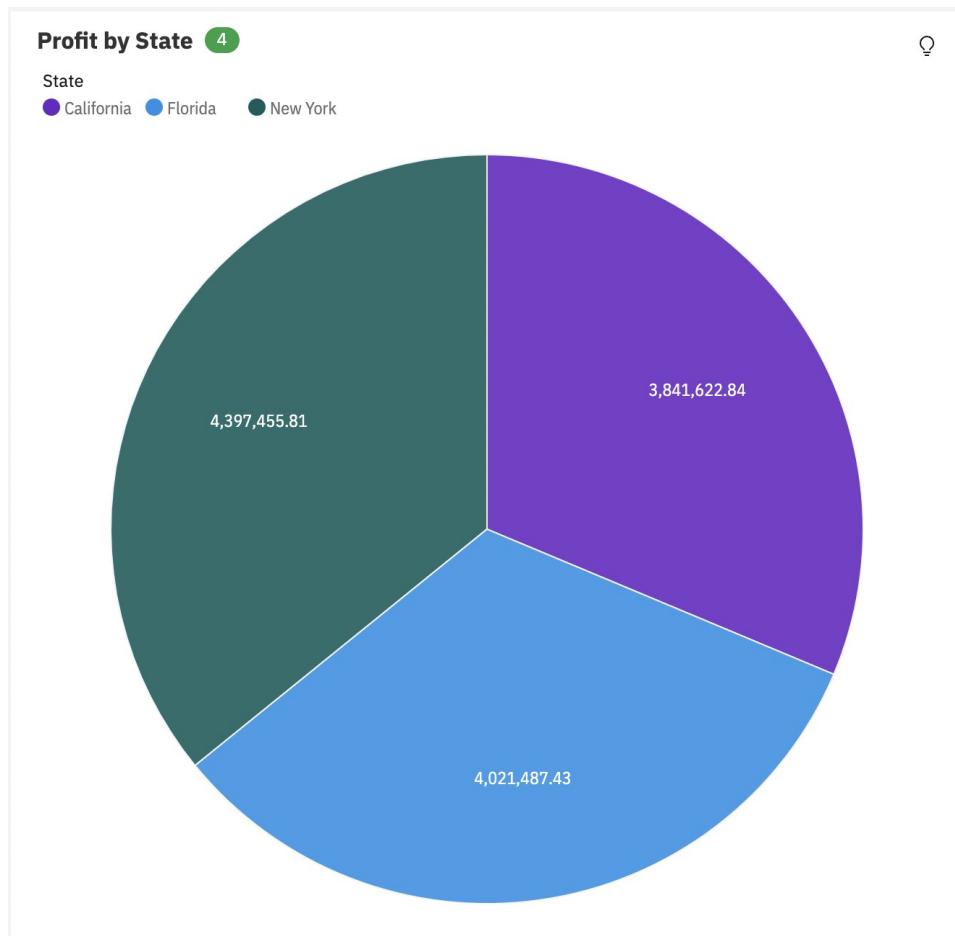
The screenshot shows the IBM Cognos Analytics with Watson interface. On the left, there's a sidebar with 'Hello. Welcome Watson.' and a 'Watch video' button. Below it is a 'Quick launch' section with a 'Get started' button. In the center, a modal window titled 'Add a data source to explore' is open. It has tabs for 'My content' (selected) and 'Team content'. Under 'My content', there's a list of uploaded files: '50_Startups.csv' (CSV), 'IPL Ball-by-Ball 2008-2020.csv' (CSV), 'IPL Matches 2008-2020.csv' (CSV), 'Olympic Events.xlsx' (XLSX), 'Olympic Medals.xlsx' (XLSX), and 'sales.csv' (CSV). The 'sales.csv' file is highlighted. At the bottom of the modal are 'Cancel' and 'Add' buttons. The background of the main interface shows a dark theme with some UI elements and a small illustration of a person at a desk.

In order to decide which visualizations to create, the data was first analyzed and the features understood. The grid data was accessed using the **Prepare Data** option.

The screenshot shows the 'Grid' view of the 'Prepare Data' interface. On the left, there's a sidebar with a 'Data module' section containing a search bar and a list of modules: 'New data module', 'Navigation paths', and '50_Startups.csv'. The '50_Startups.csv' item is selected and expanded, showing columns: '# Row Id', 'R&D Spend', 'Administration', 'Marketing Spend', 'State', and 'Profit'. The main area displays a grid of 12 rows of data corresponding to these columns. The data includes various values such as 165349.2 for R&D Spend and 192261.83 for Profit.

Row Id	R&D Spend	Administration	Marketing Spend	State	Profit
1	165349.2	136897.8	471784.1	New York	192261.83
2	162597.7	151377.59	443898.53	California	191792.06
3	153441.51	101145.55	407934.54	Florida	191050.39
4	144372.41	118671.85	383199.62	New York	182901.99
5	142107.34	91391.77	366168.42	Florida	166187.94
6	131876.9	99814.71	362861.36	New York	156991.12
7	134615.46	147198.87	127716.82	California	156122.51
8	130298.13	145530.06	323876.68	Florida	155752.6
9	120542.52	148718.95	311613.29	New York	152211.77
10	123334.88	108679.17	304981.62	California	149759.96
11	101913.08	110594.11	229160.95	Florida	146121.95
12	100671.96	91790.61	249744.55	California	144259.4

Visualization #1:



This Pie chart visualization depicts the total profit earned by startups belonging to each state present in the Dataset.

Fields

● **Segments*** Required field

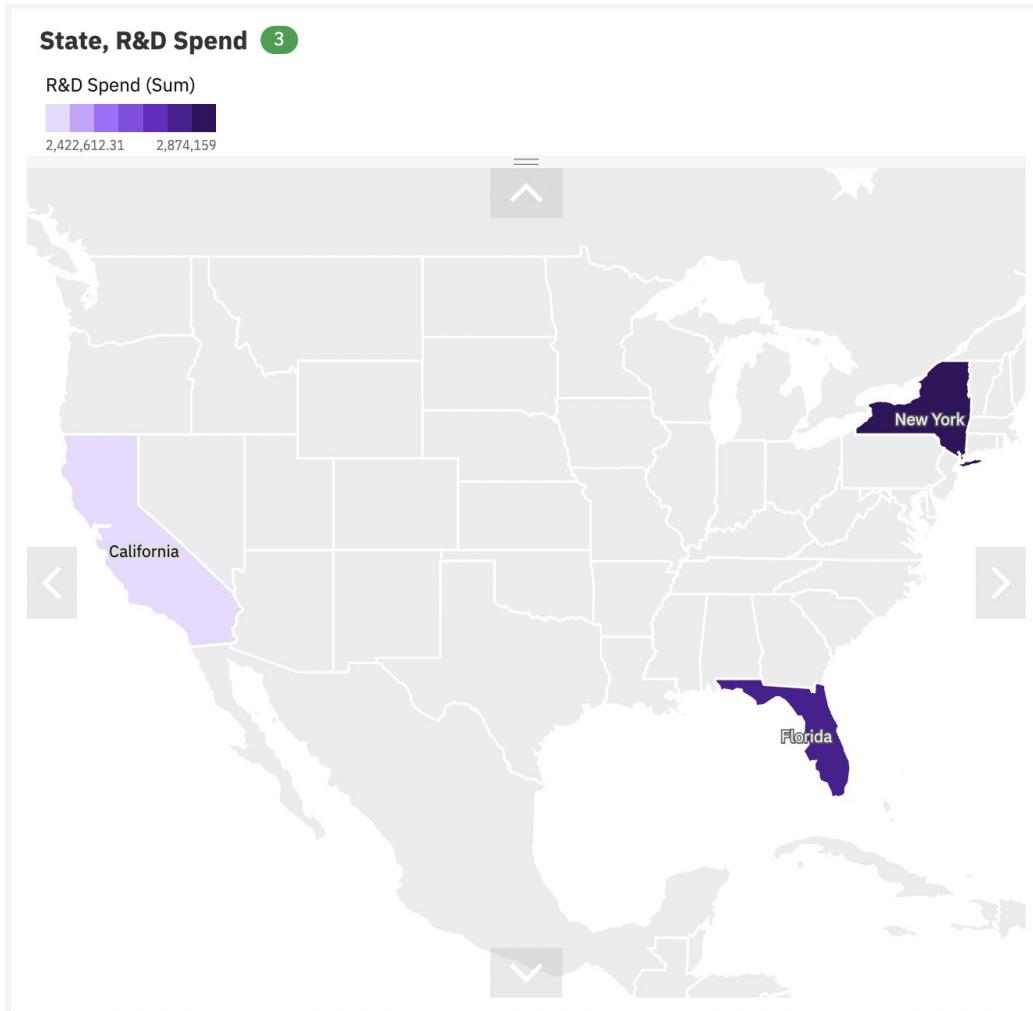
⋮ State ⋮

Click or drag data here

Size* Required field

⋮ Profit ⋮

Visualization #2:



This Geospatial Map (Geolocation) visualization shows the State-wise startup expenditure towards Research and Development field.

The color gradient indicates the amount startups in the states spent, with the startups in New York spending the most (indicated by dark purple), while California spending the least in R&D (indicated by lavender).

Fields

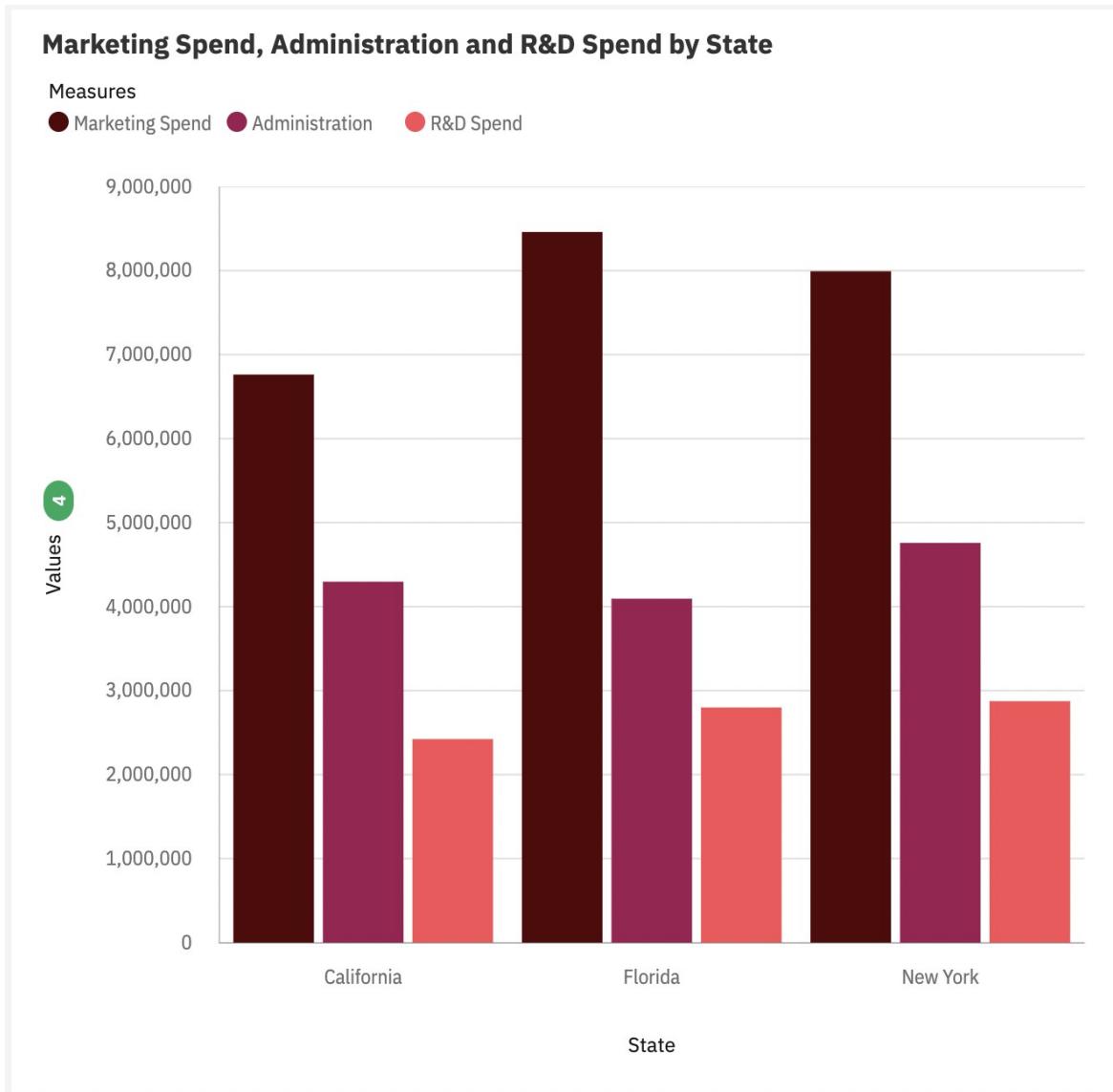
📍 **Regions*** Required field

⋮ State ⋮

■ Regions heat

⋮ R&D Spend ⋮

Visualization #3:



This visualization shows the category-wise expenditure by startups belonging to the states of California, Florida, and New York. This will instantly provide insight with regard to the amount of money spent for each category, bearing in mind the location of the startups.

Fields



Bars

⋮ State

⋮

Click or drag data here

Length*

Required field

⋮ Marketing Spend

⋮

⋮ Administration

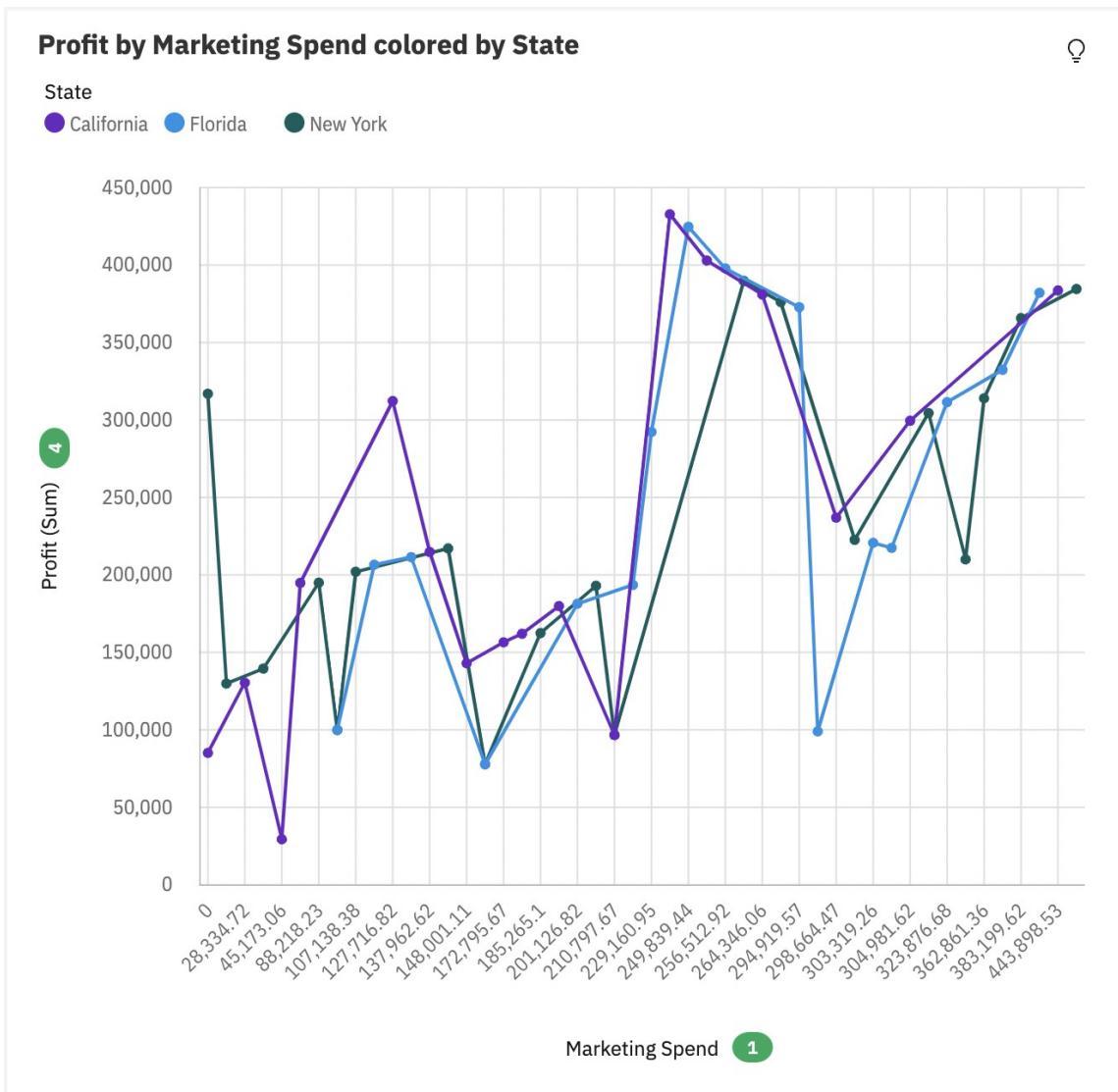
⋮

⋮ R&D Spend

⋮

Click or drag data here

Visualization #4:



This visualization provides insight regarding the profit made by comparing it with the Marketing Spend - for the startups in all three states as indicated by different colored lines.

Fields

↑ x-axis*

Required field

⋮ Marketing Spend

Click or drag data here

Color

⋮ State

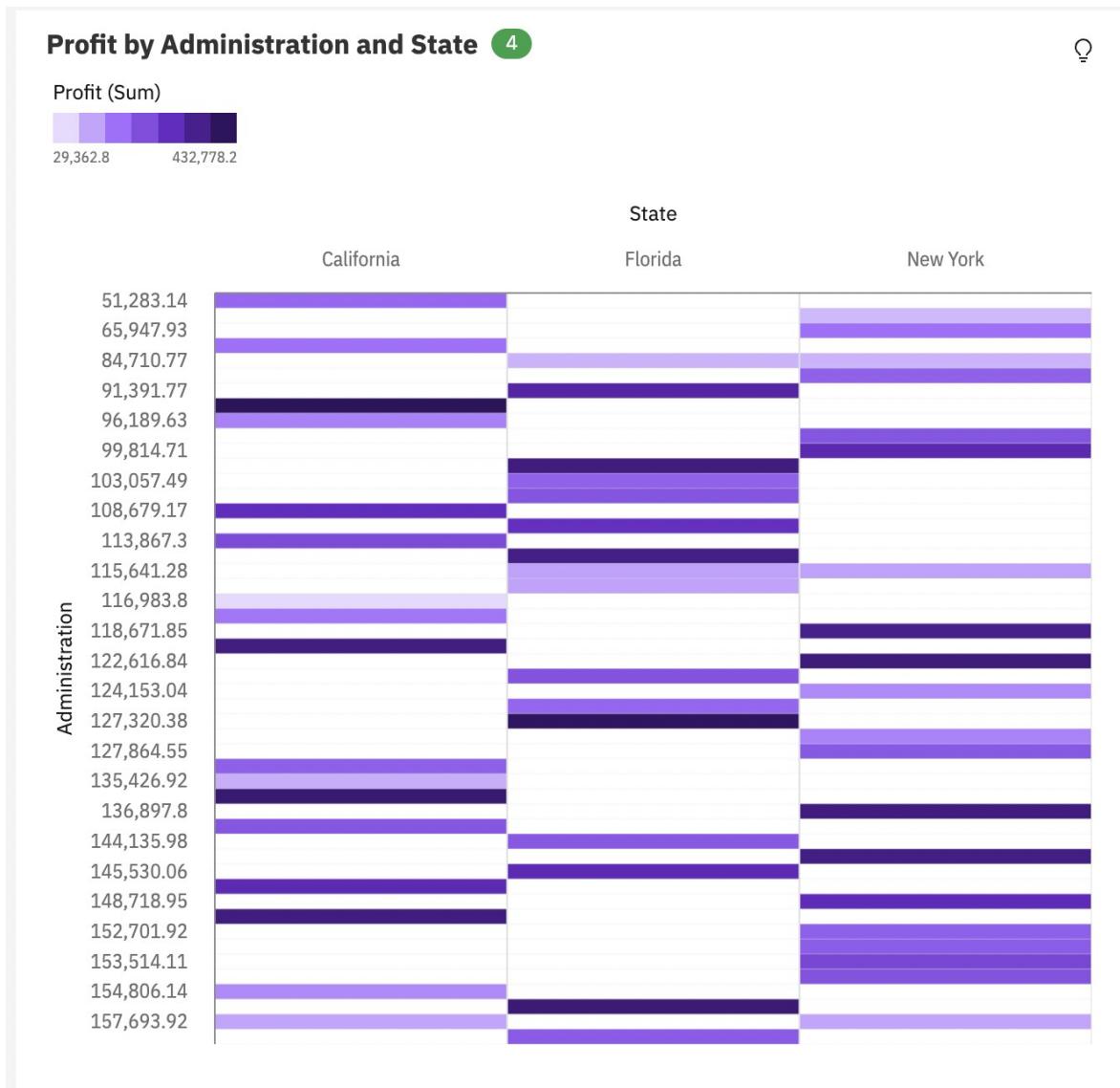
Click or drag data here

y-axis*

⋮ Profit

Click or drag data here

Visualization #5:



This visualization represents a heatmap that bases its gradient color on the profit earned - Statewise with regards to Administration spending of companies in each state.

Fields

Rows

Administration

Click or drag data here

Columns

State

Click or drag data here

Heat*

Required field

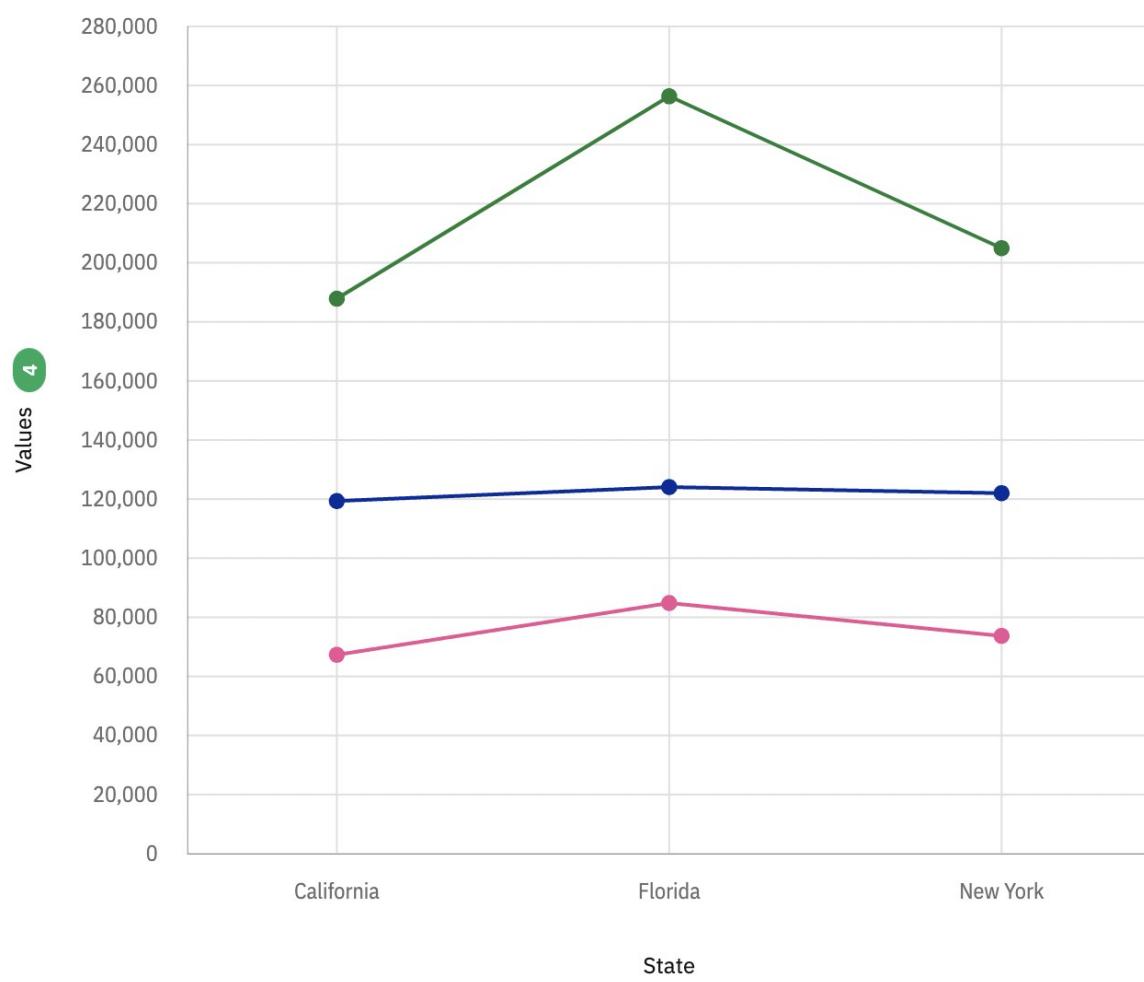
Profit

Visualization #6:

Marketing Spend, Administration and R&D Spend by State

Measures

● Marketing Spend ● Administration ● R&D Spend



This visualization represents the average count of each expenditure for a startup - Marketing spend, Administration, and R&D Spend - mapped state-wise.

Fields

↳ x-axis* Required field

⋮ State ⋮

🎨 Color

⋮ Measures group (3)

Click or drag data here

y-axis*

⋮ Marketing Spend ⋮

⋮ Administration ⋮

⋮ R&D Spend ⋮

Required field

Click or drag data here