Prompt Template:

JavaScript Copy Caption

```
Consider the below JSON array of objects describing low-level analytic tasks as a list of their “name”, “description” and “pro forma abstract" :
[
  {
    "name": "Correlation",
    "description": "Given a set of data cases and two attributes, determine useful relationships between the values of those attributes.",
    "pro forma abstract": "What is the correlation between attributes X and Y over a given set S of data cases?",
    "examples": ["Is there a correlation between carbohydrates and fat?", "Do different genders have a preferred payment method?", "What is the relationship bet
ween budget and gross?"],
    "Attributes":["Quantitative", "x" ,"Quantitative", "x" ,{"Nominal","Ordinal","Quantative","Temporal"}],
    "Attributes description":"encodings: x, y, color, size, column, row, column to visualize up to three attributes at a time. "x" indicates that the first attr
ibute is used for the horizontal or x-axis of the visualization. similarly the second "x" is used for the vertical or y-axis of the visualization. Attributes in
the curly brackets {} are optional.",
    "Visualization Recommendation":["Scatterplot"]
  },
  {
    "name": "Derived Value",
    "description": "Given a set of data cases, compute an aggregate numeric representation of those data cases.",
    "pro forma abstract": "What is the value of aggregation function F over a given set S of data cases?",
    "examples": ["What is the average calorie content of Post cereals?", "What is the gross income of all stores combined?", "How many manufacturers of cars are
there?"],
    "Attributes":["Nominal", "x" ,"Ordinal", "x" ,"Quantitative",{"Nominal","Ordinal","Quantative","Temporal"}],
    "Attributes description":"encodings: x, y, color, size, column, row, column to visualize up to three attributes at a time. "x" indicates that the first attr
ibute is used for the horizontal or x-axis of the visualization. similarly the second "x" is used for the vertical or y-axis of the visualization. Attributes in
the curly brackets {} are optional.",
    "Visualization Recommendation":["Bar Chart"]
  },
  {
    "name": "Filter",
    "description": "Given some concrete conditions on attribute values, find data cases satisfying those conditions.",
    "pro forma abstract": "Which data cases satisfy conditions {A, B, C...}?",
    "examples": ["What Kellogg's cereals have high fiber?", "What comedies have won awards?", "Which funds underperformed the SP-500?"],
    "Attributes":["Quantitative", "x" ,"Nominal", "x" ,"Ordinal",{"Nominal","Ordinal","Quantative","Temporal"},"x",{"Quantitative"}],
    "Attributes description":"encodings: x, y, color, size, column, row, column to visualize up to three attributes at a time. "x" indicates that the first attr
ibute is used for the horizontal or x-axis of the visualization. similarly the second "x" is used for the vertical or y-axis of the visualization. Attributes in
the curly brackets {} are optional.",
    "Visualization Recommendation":["Strip Plot", "Histogram", "Bar Chart", "Heatmap"]
  },
  {
    "name": "Distribution",
    "description": "Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attribute’s values over the set",
    "pro forma abstract": "What is the distribution of values of attribute A in a set S of data cases?",
    "examples": ["What is the age distribution of shoppers?", "What is the distribution of carbohydrates in cereals?"],
    "Attributes":["Temporal", "x",{"Quantitative"}, "x" ,{"Nominal","Ordinal"}],
    "Attributes description":"encodings: x, y, color, size, column, row, column to visualize up to three attributes at a time. "x" indicates that the first attr
ibute is used for the horizontal or x-axis of the visualization. similarly the second "x" is used for the vertical or y-axis of the visualization. Attributes in
the curly brackets {} are optional.",
    "Visualization Recommendation":["Line Chart"]
  }

]
```

```
Using the above definitions, classify the below natural language queries into the respective analytic tasks they map to. List the input query along with the “na
me” of the mapped tasks, Attributes name and also recommend the visualization type, add explanation to how you determined the task from the input query, not the
input query itself, also return the visualization type in the form of vegalite specification where it reads data from url:https://raw.githubusercontent.com/nl4d
v/nl4dv/master/examples/assets/data/movies-w-year.csv for vegalite specification, here is a sample dataset for reference:

Title,Worldwide Gross,Production Budget,Release Year,Content Rating,Running Time,Genre,Creative Type,Rotten Tomatoes Rating,IMDB Rating
From Dusk Till Dawn,25728961,20000000,1996,R,107,Horror,Fantasy,63,7.1
Broken Arrow,148345997,65000000,1996,R,108,Action,Contemporary Fiction,55,5.8
City Hall,20278055,40000000,1996,R,111,Drama,Contemporary Fiction,55,6.1
Happy Gilmore,38623460,10000000,1996,PG-13,92,Comedy,Contemporary Fiction,58,6.9
Fargo,51204567,7000000,1996,R,87,Thriller,Contemporary Fiction,94,8.3

    Query can have more than one task and visualization type they can map to:

Show how content ratings are associated with gross receipts by genre.
```

Response:

Query: "Show how films' Rotten Tomato ratings compare with IMDB ratings."

Mapped Tasks:

1. Task Name: Correlation
   - Attributes: Quantitative (Rotten Tomatoes Rating), Quantitative (IMDB Rating)
   - Visualization Recommendation: Scatterplot
   - Explanation: The query aims to compare the Rotten Tomatoes ratings with the IMDB ratings of films. It involves determining the correlation between the quantitative attributes "Rotten Tomatoes Rating" and "IMDB Rating" to understand the relationship between these two variables.
2. Task Name: Distribution
   - Attributes: Quantitative (Rotten Tomatoes Rating), Quantitative (IMDB Rating)
   - Visualization Recommendation: Line Chart
   - Explanation: The query also seeks to explore the distribution of Rotten Tomatoes ratings and IMDB ratings separately. By visualizing the distribution of each rating using a line chart, we can observe the frequency or density of different rating values.

Visualization Specifications (Vega-Lite):
For Correlation:

json Copy code

```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v4.json",
  "data": {"url": "https://raw.githubusercontent.com/nl4dv/nl4dv/master/exam
  "mark": {"type": "circle", "tooltip": true},
  "encoding": {
    "x": {"field": "Rotten Tomatoes Rating", "type": "quantitative"},
    "y": {"field": "IMDB Rating", "type": "quantitative"}
  }
}
```

For Distribution (Rotten Tomatoes Rating):

json Copy code

```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v4.json",
  "data": {"url": "https://raw.githubusercontent.com/nl4dv/nl4dv/master/exam
  "mark": "line",
  "encoding": {
    "x": {"field": "Rotten Tomatoes Rating", "type": "quantitative", "bin":
    "y": {"aggregate": "count", "type": "quantitative"}
  }
}
```