About the dataset:

The dataset provided to us namely ‘employment-and-wages-annual-data-beginning-2000’ is a quarterly census data that encompasses information about employees and their wages at the various area types such as state, county, labor market area, metropolitan statistical area and workforce investment region. The dataset has 9 features namely Area Type, Area, NAICS (code given to various work), NAICS title (Name of the industry/domain), Year, Establishment (Number of organizations in the domain), Average Employment, Total Wages and Annual Average salary. Out of the 9 given features 3 features Area Type, Area and NAICS are categorical and the rest of the 6 features are numerical.

Table

Description automatically generated

Fig. 1 Screenshot of dataset

Use of dataset:

This dataset can be used to study detailed statistics about various industries (also can be used to get an idea about which industry employs maximum or minimum depending on their location i.e., area type), their employment depending. Also, this dataset can be used to visualize annual salary an industry based on their location. Plus, the dataset is also helpful in getting a general idea about the total wages and number of establishments in a particular job domain.

Chart, scatter chart

Description automatically generated

Fig. 2 Total Wages V/S Average Employment

Inference from visualization:

In this scatterplot we created 3 categories: High, Medium and Low depending on their position in the scatter plot (Wages vs Average Employment)

This graph shows us which areas have high average employment and total wages (New York City and New York State), areas which have medium average employment and total wages (New York County, Balance of state and New York City Region) and areas with low average employment and total wages (Long Island Region, Hudson Valley region, etc.)

Filter has been applied to the years measure to display the data of the year 2016 and on total wages as it consisted of 7 null values which needed to be taken care of.

The aim of creating this sheet was to create a scatterplot to visualize the "Total Wage" versus "Average Employment".

Further, in the plot, shows that New York City (area) having the average employment of 32,184,820 has the Highest Total Wage in 2016.

Also, as required by the question the marks display the Annual Average Salary in that area i.e., 18,024,872.

Chart, scatter chart

Description automatically generated

Fig. 3 Total Wages V/S Average Employment (All Private Sector)

Inference from visualization:

The above scatter plot was plotted to find out why there is a difference between the employment and total wages of New York City and New York State.

To analyze that we further divided the category Area into 2 parts: One containing NAICS title (All private: Sheet 2) and another NAICS title(All Government) to evaluate if the employment by these two sector is due to which there is a gap in the average employment and total wages of both the regions.

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

The comparison between the Private sector yields the following result:

|  |  |  |
| --- | --- | --- |
| Region | Average Employment | Total Wages |
| New York City | 7,253,794 | 646,354,549,012 |
| New York State | 7,781,866 | 537,782,552,480 |

Chart, scatter chart

Description automatically generated

Fig. 4 Total Wages V/S Average Employment (All Government Sector)

Inference from visualization:

Having further analyzed the private sector we move forward and check whether the employment and wages of the government sector are the reason for disparity for the two regions.

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

The comparison between the government sector of the two regions show us that the average employment by the New York State is higher but the total wages of the New York City is much more. This difference is due to the total wages earned by the private sector of New York City is very much greater than that of New York State.

Chart

Description automatically generated

Fig. 5 Comparison of Annual Average Salary and Establishments (Area Wise)

Inference from the visualization:

This bar plot helps us infer why the average employment in New York State is more.

The reason being the number of establishments present in the region is more. This trend follows for the whole dataset as the regions which have a higher number of establishments generate more employment.

But that does not necessarily mean that the annual average salary of the region is more which can be clearly seen in the graph above.

New York City has a significantly smaller number of establishments then New York State but the average annual salary for New York City is comparatively greater. The same inference can also be made about New York County and New York City Region.