**NIC-DRIVEN INDUSTRY METRICS: ASI DATA EXPLORATION**

1. **A Brief introduction about the dataset**

The dataset that I have chosen is the **Annual Survey of Industries (ASI)**. The Annual Survey of Industries (ASI) is conducted by the National Statistical Office (NSO) in India, providing comprehensive data on the industrial sector. This survey primarily focuses on large-scale manufacturing establishments across the country, and the data is categorised according to the **National Industrial Classification (NIC)**. The NIC is a standardised system used in India to classify and categorise economic activities for statistical purposes. ASI covers all registered factories in India under the Factories Act of 1948 and includes manufacturing, repair services, and other industries. Each establishment is classified by NIC code, which helps in grouping similar industrial activities and comparing them on a national and regional level. The survey gathers data on numerous economic variables, such as employment, wages, productivity, capital investment, output, input costs, and value addition across industries. It also covers information related to fixed assets, working capital, and depreciation, enabling analysis of profitability and efficiency. NIC codes are used to systematically classify industries, allowing data users to identify patterns and trends within specific industry categories. The ASI reports data according to NIC-2008, allowing industries to be examined by sector, such as manufacturing, utilities, and repair services. ASI data is crucial for understanding the industrial sector's structure, growth, and composition. It aids policymakers in identifying growth areas and allocating resources efficiently. The data also helps researchers and economists track the performance of specific industries and assess the impact of industrial policies. The ASI is an annual survey, and its data is made available to the public through government publications and online databases. ASI data, classified under NIC codes, is widely used for economic analysis, industrial planning, and employment studies. The variables used in the dataset are as follows: Sector, Type of organization, Type of ownership, Whether the unit has ISO Certification, Input, Output, GVA, NVA, Depreciation, Net Income, Profit, Average Salary, Labour, Bonus, PF, Capital, Wages, etc. he classification of industries includes Sections(A-U), Divisions (99), Groups (990).

1. **Variables used and their scale of measurements used for visualisation**

**1. Input (in million rupees)**- Purchase value of total (indigenous and imported) input items consumed + Expenses of work done by others on materials supplied by other units + repair & maintenance + operating expenses + insurance charges + rent paid for pants & machinery and other fixed assets. **2. Output (in million rupees)** (Sale value of the total of products and by-products manufactured by the unit) – (excise duty, sales tax, other taxes) + (Working capital and loans on finished goods) + (receipts from electricity generated and sold, value of own construction, net balance of goods sold in the same condition as purchased, rent received from plants and machinery, income from services, semi-finished goods). **3. GVA (in million rupees)**= Output – Input. **4. Type of sector**- Urban and Rural. **5. Type of Ownership** includes- a) Wholly Central Government, b) Wholly State and/or Local Govt. c) Central Government and State and/or Local Government jointly, d) Joint Sector Public, e) Joint Sector Private, f) Wholly Private Ownership. **6.** To calculate no. of factories, **the count of DSL** is used. **7. Average Salary (in rupees)** = Total salary divided by number of labourers. **8. Wages (in million rupees) =** Remuneration as related to an individual worker, in terms of money, directly or indirectly payable, more or less regularly for each pay period, in respect of his/her employment or work done in such employment. In the visualisation, the Measure value Average of wages is used. **9. Profit (in million rupees) =** (Net Income – Wages – Bonus – PF – Welfare). Average Profit is used. **10. Welfare(in thousands rupees)** = Workmen & staff welfare expenses (in Rs): These include expenditure incurred by the employer on maternity benefits and crèches and other benefits such as the supply of food, beverages, tobacco, clothing and group lodging at confessional rates and educational, cultural and recreational facilities and services and grants to trade unions and cooperative stores meant for employees. All group benefits are included. Average Welfare used.

**GROUPING**- states are grouped into **regions** using **State code** – Northern Region, Western Region, Southern Region, Eastern Region, and Central Region. Industries are grouped into different types according to **NIC2digit data**- Manufacturing, wholesale and retail trade, Agriculture, forestry and fishing, Electricity, gas, steam and air conditioning supply, Transportation and storage, Water supply, Information and communication, other service activities, Mining and Quarrying, Administrative and support service activities, Professional, scientific and technical activities, construction and Accommodation and Food service activities.

**CALCULATED FIELD**- A calculated field named **Perks (in million rupees)** is created which includes [Wages + Welfare + Emoluments + Bonus + PF]. The average of this field is used for visualisation.

**PARAMETER**- A parameter named **TOP N** is created which shows the top N Industries based on Average Perks.

|  |  |
| --- | --- |
| **Variables** | **Scale of Measurement** |
| Input | Million Rupees |
| Output | Million Rupees |
| GVA | Million Rupees |
| Average Salary | Rupees |
| Wages | Million Rupees |
| Profit | Million Rupees |
| Welfare | Thousand Rupees |
| Perks | Million Rupees |

1. **PROCESS OF MAKING DASHBOARDS**

**Understanding the data**

The dataset is the annual survey of industries, which is conducted by the Statistical Office of India. It provides comprehensive data on the industrial sector, and the data is categorised according to the National Industrial Classification 2008. It contains data from the years 2010-2011 and 2013-2014. The goal is to allow data users to identify patterns and trends within specific industry categories and also track developments in industries, which can help policymakers formulate policies for the future.

**Importing and Preprocessing data**

By drag and drop, we connect the data into tableau. There are 5 sheets in the ASI .xlsx file, so we connect each one of them. Then, we proceed to the worksheet. Before making visualisations, We map the names of states according to state codes to have a better understanding. Grouping of Industries into different types of industries is done based on different NIC2digit codes. Renaming of type of sector from 1 and 2 to Rural and Urban. A group called Region is created to group states. Alias of Types of Ownership is changed to have an easy understanding. Some null values are hidden from the data source.

**Building the visualizations**

The variables used in the dataset are as follows: Sector, Type of organization, Type of ownership, Whether the unit has ISO Certification, Input, Output, GVA, NVA, Depreciation, Net Income, Profit, Average Salary, Labour, Bonus, PF, Capital, Wages, etc. There are seven visualizations on the dashboard. The first one is a map that shows no. of factories state-wise and region-wise with a colour gradient of Red and Gold where lighter colour shows a lesser value and vice versa. There are 2 bar plots – one shows the total input, total output, and GVA of Rural and Urban sectors, and the other one shows the relationship between Average perks and Type of industry on the y-axis and x-axis, respectively. It also has an average line passing through the bars. The fourth visualization is a line graph showing the trends in average salary industry-wise. The Y-axis has the average salary, and the X-axis has the type of industry. The fifth visualization is a whisker plot representing the average profit by ownership. The colour gradient shows highest to lowest in darker to lighter tones. The sixth is a scatter plot between wages and welfare in different industries. The last visualization is a tree map showing the average salary region-wise with lighter to darker tones from lowest to highest.

**Creating Dashboard and Formatting**

Here, I used tiled method of visualizations with a dashboard title named “NIC-Driven Industry Metrics: ASI Data Exploration”. The visualizations have a dark red border, which makes them distinct from each other. Titles are coloured black. There is a constant beige background. The legends like average salary, average profit, and perks are on the right side of the dashboard pane. The colour gradient of red and gold is evident in the legends There are four filters, all synced on the data source so that there is interactivity between different visualisations. The tiles of legends and filters have an orange border with a darker shade, making them distinct from visualizations.

**Filters and Interactivity**

Four filters are used on the dashboard- Sector, Ownership, Regions, industry type, and Top N based on Average perks. All are connected to the data source and work on all sheets and visualisations.

1. **Description of Visualization and its analysis**
2. **Input, Output and GVA**

This is a bar graph having input, output, and GVA on X- the X-axis distributed in Rural and Urban sectors. The colour gradient chosen is red- gold. The dark red shows the large value, and the yellowish bar shows the lower side of the magnitude. In the rows, measure values like the sum of input, the sum of output, and the sum of GVA are used. In the column section, types of sectors is used. Different filters used are like ownership, sector, industry, and regions. At first glance, the data in both sectors are quite comparative.

On digging deep and filtering based on the type of ownership, we got to know that under Central govt ownership, under the rural sector, all three parameters – input (158 billion), output (197 billion), and GVA (39 billion) are quite low as compared to urban sector-input (977 billion), output (1140 billion) and GVA (163 billion). Under state govt ownership, the urban sector is not as evident as the previous one. The joint sector public trend is the same as the central, with the urban sector having more chunks. For the joint sector private, the rural sector’s output (595 billion) is more than the urban sector’s (483 billion). In private ownership, both input and output are comparable in both sectors.

When filtering on regions, the Southern region has the highest values. In the western region, the Rural sector data (input=16920 billion, output=20770 billion, GVA=3847 billion) is more evident than the urban sector (input=10424 billion, output=12636, GVA=2212 billion). For the electricity, gas, steam type of industry, the rural output is 1593 billion and urban output is clocked at 196 billion. Wholesale and retail industry has the rural parameters (input=4 billion, output= 10 billion and GVA=6 billion) and the urban parameters at (input=45 billion, output=95 billion and GVA= 50 billion) which is quite contrasting. Agriculture and manufacturing have quite comparable inputs and outputs in both the rural and urban sectors. The policies should be introduced to minimise the gaps where it is evident.

1. **No. of Factories State-wise and Region-wise**

This visualisation is a choropleth map of India, displaying the number of factories across different states and regions. The map uses varying shades of orange to represent factory counts, with darker shades indicating higher numbers. The colour gradient, shown in the legend on the right, ranges from light orange for states with fewer factories to dark brown for states with the highest counts.

Key insights from this visualisation include the concentration of factories in certain states, particularly Maharashtra, Gujarat, and Tamil Nadu, which are shaded in the darkest tones and show counts of over 10,000 factories each. This highlights these states as industrial hubs, possibly due to better infrastructure, resource access, and more developed supply chains. In contrast, states in northeastern and some northern regions, such as Sikkim, Mizoram, and Manipur, have much lower counts, reflected by lighter shades on the map, suggesting limited industrial activity. These variations can be attributed to factors like geographic constraints, lower levels of industrialization, or economic policies that do not favour heavy industrial growth. The highest number of factories is in Tamil Nadu, with 10,477 factories, followed by Maharashtra, with 10,014. For the lowest one, Andaman and Nicobar has 25 factories. The majority of the factories are in the urban sector. The number of factories under central govt ownership is quite low. The same is the case with state govt ownership. The majority is in private and not-defined categories, with not-defined being the highest. Not defined probably means the ownership is not clear cut recognised in either category. Southern region states have the highest number of factories. The western region states like Maharashtra (10000), Gujarat (7500), and Gujarat (3500) also have large chunks of factories, maybe because of the liberal laws there. The manufacturing industry has the highest number, and construction is the lowest. Bihar has no agricultural-related factory. As far as Uttar Pradesh is concerned, the transportation type factories are the highest. Factories involved in water supply industries are the highest in Gujarat. Maharashtra (158) and Tamil Nadu (150) are agriculture-type factories.

1. **Average Salary Region-wise**

This visualisation is a Treemap representing the average salary by region across various parts of India. Each coloured rectangle corresponds to a region, with the size and colour intensity reflecting the average salary level within that region. Darker shades indicate higher average salaries, while lighter shades represent lower ones. The colour gradient ranges from dark red to yellow, as shown in the legend on the right, with values spanning from approximately ₹243.61 to ₹330.19.

In this Treemap, the Western Region has the highest average salary, represented by the largest and darkest red block, suggesting that it stands out among the regions regarding salary levels. The Northern and Central Regions, with moderately intense orange shades, follow with slightly lower average salaries than the Western Region but are still higher than the rest. Finally, the Eastern Region, with the lightest yellow shade, has the lowest average salary, as indicated by its smaller area and lighter colour tone.

When filtering on the sector, in the rural sector, the southern region provides the highest salary (325.8), followed by the western region (318.8), and the lowest is the eastern region (208.5). the urban sector, a western region, with a salary of 338, is the highest, followed by the southern (309), and the lowest among them is the central region (249). The northern region shows the highest average salary in centrally owned factories. The western region has the highest average salary among state-owned factories. In the case of privately owned factories, the western region shows the maximum value

This visualisation highlights the significant regional disparities in average salary across India. The Western Region, with the highest salary levels, likely benefits from a combination of factors such as industrialisation, infrastructure, and economic activity, contributing to higher wages. As part of the Western Region, Maharashtra and Gujarat are industrial and commercial hubs, possibly leading to a higher demand for skilled labour and better wages. The Eastern Region's low average salary may reflect its relatively underdeveloped industrial base, limited employment opportunities in high-paying sectors, and a larger informal workforce. These factors often contribute to lower wage levels in the region, and this trend is clearly captured by the treemap’s colour coding and sizing. One advantage of this treemap is that it clearly compares average salaries by region in a single view. The colour and size cues enable easy identification of the regions with the highest and lowest salary levels.

1. **Wages and Welfare comparison industry-wise**

This is a scatter plot visualization that signifies the relationship between wages and welfare industry-wise. Here, the variables used are wages, welfare, and NIC2digit groups. For measure values, the average of wages and welfare are used. Red colour signifies average wages and orange colour shows average welfare. the filters used are industry type, sector, and ownership.

The highest welfare is found in the Manufacturing industry, with a value of 3.37 million, followed by electricity, gas, and steam, having a 3.08 million average welfare. The accommodation and food service industry has the lowest welfare value of 70,091. In the case of average wage, the construction industry tops the chart with a staggering 91 million, and the accommodation and related industries are lowest at 4.6 million. By adding a sector filter, we find that the construction industry has the highest wages in the rural sector, and information and communication have the highest wages in the urban sector.

Upon digging deep through ownership filtering, we find out that in centrally controlled industries, the manufacturing industry has the highest welfare (65 million) and wage (101 million). In state-controlled industries, the same manufacturing sector has the highest wage and welfare (32 million and 2.9 million, respectively), followed by the electricity, steam, and gas industries. The other services industry category has the highest wage of 72 million, and information and communication have the highest welfare of 2.4 million.

Regarding regional segregation, we find that in the central region, the electricity industry has the highest wage (41.6 million), followed by manufacturing, and the lowest is transportation (6.49 million). As far as welfare is concerned, the electricity, gas, and steam industries (3.6 million) are the highest, followed by the manufacturing industry (3.50 million). In the eastern region, information and communication provide the highest average wage of 42.7 million, and welfare is highest in the electricity, steam, and gas industries (3.20 million). In northern states, the highest average wage is among construction factories (344 million), and welfare is highest in the information and communication industry (5.6 million). The transportation industry provides the highest average wage in the southern region (90.4 million), and welfare is highest in manufacturing (3.9 million). In the western region, administrative and support service activities provide the highest average wage of 47.09 million. As far as welfare is concerned, the electricity, gas, and steam industries have the highest average welfare in this region (4.36 million).

1. **Average profit ownership-wise**

This visualisation uses the whisker plot to show the average profit among factories divided by ownership. Here, the colour gradient is the same red-gold as above, with darker shades representing higher values and lighter shades representing lower values. In the mark card, colour is chosen by average profit, and the label is of ownership. The filters used are type of industry, region and sector.

At first glance, we can see that the private joint sector has the highest average profit among all, with a value of 1116 million. In the 2nd position, centrally owned industries come with an average profit of 649 million, followed by joint sector public-owned industries with a 588 million average profit in the 3rd position. Then, a large gap shows the high disparity in average profit on distant 4th, centrally and state government jointly owned industries earn an average profit of 91 million followed closely by industries whose ownership is not well defined with an average profit of 80.2 million and privately owned industries with an average profit of 74.7 million. State government-owned industries have a negative average profit, meaning a loss of 1.8 million.

On further filtering on sectors, we find out that in the rural sector, the list is the same as the whole, with joint sector private as an outlier with an average profit of 1943 million. The state govt. has the lowest average profit of 294,450. In the urban sector, the joint sector public has an average profit of 1114 million, followed by centrally owned industries with an average profit of 783 million. On the lower side, both state-owned industries (-3.3 million) and industries with central and state joint ownership have a loss of 46 million.

On the region segregation, we get more information like in the central region, centrally owned industries are an outlier with an average profit of 4115 million, and the lowest is joint sector private with an average loss of 780 million. In the eastern region, the joint sector public has the highest share of average profit with a value of 722 million, followed closely by joint sector private with 700 million. In this region, state government-owned industries have an average loss of 56 million. In the northern region, central and state govt joint industries' average profit is the highest, with a value of 1044 million. Here, central government-owned industries have an average loss of 36.7 million. In the southern side of the country, the joint sector public-owned industries have the highest average profit of 808 million, followed by centrally owned industries with an average profit of 548 million. An average loss of 83.4 million is seen in central and state govt jointly owned industries. Lastly, in the western region, the private sector has an average profit of 2945 million, and on the negative side, we have centrally and state government-owned joint industries with a loss of 228 million.

1. **Average perks**

This Tableau visualization is titled "Average Perks," which shows the average perks across different types of industries. The graph is a bar chart, where each bar represents a type of industry. Key fields like "NIC 2-digit (group)," "Region," "Sector," and "Ownership" is applied as filters, indicating that the data is segmented based on these categories. A calculated field named **Perks** is created, including [Wages + Welfare + Emoluments + Bonus + PF]. The average of this field is used for visualization. The X-axis has types of industries. The y-axis shows the average perks in millions.

The red bar, which towers over the others, reaches around 245.2 million, making it a significant outlier in the dataset. The average value marked around 61 million (indicated by a dotted red line. The construction industry with the highest perks (245.2 million) significantly deviates from the others, suggesting either a highly lucrative field or perhaps one with a smaller employee base, thus skewing the average. Industries with perks above average are construction (245.2 million), information and communication (119), manufacturing (77 million), electricity, steam and gas (76 million).

In the rural sector, administrative activities services have the highest perks of 124 million, and the lowest is agriculture (16 million). In the urban sector, the information and communication industries have the highest perks, with a value of 129 million, and accommodation and food services activities have the lowest average perks of 11 million. In privately owned industries, information and communication industries provide the highest perks of 107 million, followed by other services with 103 million average perks. Among centrally owned industries, electricity, steam and gas industries have the highest average perks of 80 million.

In the central region, manufacturing has the highest value of 86.5 million. Info and communication industries provide the highest perks in the eastern region (102.6 million). The transportation industry has the highest average perks in the northern region, with a value of 107 million. In the southern region, information and communication have the highest average perks of 108 million. In the Western region, most industries have comparative average perks and are well above average, which suggests better working products in Western states. The manufacturing industry provides the highest average perks in the Western region (88 million). This data can guide professionals or job seekers in choosing sectors that offer higher perks. For companies, understanding where they stand in terms of perks relative to other sectors can help adjust their packages to attract talent.

1. **Average Salary Industry-wise**

This visualization titled "Average Salary Industry-wise" shows a trend in average salaries across different industries, represented by a line chart. The chart offers a clearer view of how salaries deviate in different industries, indicating that some industries consistently offer higher compensation than others. The X-axis shows the types of industries. The y-axis represents average salaries. Key fields like "NIC 2-digit (group)," "Region," "Sector," and "Ownership" are applied as filters.

The trend starts with the highest salary at 501.3, declining steadily to the lowest salary point at 201.9. The highest-paying industries have salaries around 501.3 and 497.3, indicating that certain industries provide significantly higher average compensation. Lower salaries appear in industries with averages like 225.9 and 201.9, highlighting sectors that may rely more on lower-paid, labour-intensive roles or those with large workforces that drive down the average salary. The Information and communication industry has the highest average salary, valued at 501, followed by electricity, steam and gas (497). On the lower side, agriculture has an average salary of 225; the lowest is accommodation (201).

In the rural sector, professional scientific activities provide the highest salary of 545.7 and the lowest is the agriculture industry, with a value of 240.6. In the urban sector, construction and information and communication industries are in the top two, with values of 521 and 518, respectively. Accommodation and food-related services mark the lowest average salary of 202. Regarding ownership, among centrally owned industries, the water supply group has the highest average salary (724). In state-owned industries, the electricity, gas, and steam industry clocks the highest value of 693, with agriculture and allied industries being the lowest (149). Privately owned industries, information, and communication have the highest average salary of 378.4.

In the central region, the mining and quarrying industry shows an abnormally very high average value of 6162, which is an outlier. It is followed by the electricity industry (378.4). construction-based industries provide the highest salary of 537.5 in the eastern region. In the northern part of the country, electricity is the highest (712) and construction is the lowest (205). In the southern region, professional and scientific activities have the best figures of 532 and accommodation the least (201). Information and communication and agriculture-based industries have the highest and lowest average salary in the western region with the values 553 and 217, respectively. This information is valuable for job seekers assessing potential salary expectations across sectors. For businesses, this data can serve as a benchmark to understand where their industry stands regarding salary offerings and may inform adjustments to remain competitive.