# Task 01

**Expectation of the Analysis and Benefits in Informed Decisions Making Related To ministry of industry and commerce of Sri Lanka Based Problem Solving.**

The status of a job is a major factor in determining policies and tactics in the modern world of national development planning. Comprehending the elements that contribute to an occupation's renown is essential for efficient human resource management. With an emphasis on the Ministry of Industry and Commerce in Sri Lanka, we explore the use of data science in revealing insights into the variables determining occupation prestige in this paper.

**The Importance of Occupation Prestige:**

A strong toolkit to analyze big datasets and deriving useful insights is provided by data science. Utilizing cutting-edge analytics methods like predictive modeling and machine learning algorithms, we may find trends and correlations in the data that could affect the reputation of a given occupation.

**Factors Contributing to Occupation Prestige:**

1. Education: A profession's prestige can be greatly impacted by the degree of education needed for it. Jobs requiring a great deal of experience and training are usually more respected.

2.Income: The perceived prestige of different vocations can be influenced by differences in income. In society, high-paying occupations frequently enjoy more respect and distinction.

3.Gender Representation: The status of women can be impacted by their presence in particular professions. As gender dynamics change, traditionally male-dominated fields may be viewed differently.

4.Census Information: The demographic patterns seen in census information can reveal how society views certain professions and how this affects their status.

5.Occupation Type: A person's perception of a career can be influenced by the type of job they do, such as manual labor or intellectual interests.

**Advantages of Data Science for Decision-Making:**

1.Precision Planning: Through the examination of past data and patterns, policymakers can more precisely project future labor market demands, enabling more focused human resource planning.   
2. Identifying Trends: By using data science to find new patterns in occupation prestige, policies and strategies can be proactively adjusted to reflect changing public attitudes.  
3. Optimizing Resource Allocation: By directing expenditures toward sectors and occupations with the greatest potential for expansion and advancement, data analysis insights can help make judgments about resource allocation.

4.Mitigating Bias: Data-driven methods aid in reducing the biases present in human decision-making, fostering justice and equity in the allocation of resources among various industries and professions.   
5. Strengthening Competitiveness: By utilizing data science, governments can create an atmosphere that draws and keeps elite people, giving them a competitive advantage in the global economy.

The Sri Lankan Ministry of Industry and Commerce's human resource planning and decision-making procedures stand to benefit greatly from the application of data science. Through the rigorous examination of data, policymakers may identify the factors that contribute to occupation prestige and establish solutions that advance social fairness, economic prosperity, and national development. Adopting data science as a fundamental component of decision-making enables policymakers to confidently and anticipatorily negotiate the intricacies of the contemporary labor market.

# Task 2

**Tools, Techniques and Methodologies**

**Tools**

Understanding the elements influencing an occupation's prestige is crucial for national development planning. This report describes the methods and tools to be used for analyzing occupation prestige in Canada as part of the Ministry of Industry and Commerce's effort to improve decision-making using data-driven insights.

**Data Science Tools for Analysis:**

Software for Statistical Analysis (SAS): SAS is an effective tool for carrying out in-depth statistical analysis on big datasets. SAS will play a key role in finding connections and trends in the occupation dataset thanks to its powerful modeling and data manipulation capabilities.  
  
Python Programming Language: A multitude of modules and packages tailored to data analysis and machine learning activities are available in Python. Python will make data preprocessing, exploratory data analysis, and predictive modeling easier by utilizing libraries like Pandas, NumPy, and Scikit-learn.

Another flexible programming language that's frequently used for statistical computing and data visualization is called R. R will work in conjunction with Python to explore intricate linkages within the employment information and produce perceptive visuals thanks to its extensive ecosystem of data analysis packages.

Machine Learning methods: To find trends and prediction insights about occupation prestige, machine learning methods including regression, classification, and clustering will be used. While clustering algorithms can divide up the workforce based on occupations with comparable prestige traits, algorithms such as linear regression can be used to determine how education and wealth affect prestige scores.

Natural Language Processing (NLP): To extract pertinent information about the prestige of occupations, NLP techniques will be used to evaluate textual data, such as job descriptions or census reports. Text mining algorithms will make it possible to identify important themes and feelings connected to various professions, which will add to a comprehensive picture of their status.  
  
Data Visualization Tools: Effectively communicating insights to stakeholders requires the ability to visually represent the outcomes of data research. The occupation information will be used to build informative charts, graphs, and interactive dashboards that illustrate trends and interactions using programs like Matplotlib and Seaborn in Python and ggplot2 in R.

**Methodology:**

Data Preprocessing: Handle missing values, standardize data formats, and encode categorical variables to clean and prepare the occupation dataset. This guarantees that the data is prepared for modeling and analysis.  
  
Investigate the distribution of profession prestige ratings and the connections between other variables, such as gender representation, income, and education, by conducting exploratory data analysis, or EDA. For additional analysis, EDA techniques can be used to find trends, outliers, and possible areas of interest.

Statistical Analysis: To measure the links between factors and their influence on occupation prestige, run statistical tests such as correlation analysis and hypothesis testing. This stage will give empirical proof of the elements influencing the prestige of a profession in Canada.

Machine Learning Modeling: Using machine learning methods, create predictive models to estimate profession prestige scores based on pertinent attributes. Metrics such as accuracy, precision, and recall will be used to evaluate the performance of the models in order to determine how well they capture the complexity of occupation prestige dynamics.  
  
  
Text mining and NLP analysis: Use NLP methods to glean information about vocations from textual materials like job descriptions or census reports. The analysis will be further contextualized by text mining algorithms, which will assist in identifying keywords, themes, and feelings linked to high-prestige professions.

Data Visualization and Interpretation: To help stakeholders understand and make decisions, present the findings of data analysis using interactive dashboards, charts, and graphs. Clear and succinct communication of the analysis's main conclusions and insights is necessary to promote the development of evidence-based policy.

Through the application of a variety of data science techniques and technologies, the Ministry of Industry and Commerce may learn a great deal about the elements that influence the prestige of occupations in Canada. These methods, which range from statistical analysis to text mining and machine learning modeling, provide a thorough method for comprehending the dynamics of occupation prestige and guiding strategic choices for national development planning.

**Techniques**   
**Techniques for Occupation Prestige Analysis:**

The occupation dataset's salient features will be enumerated using descriptive statistics, which include measures of central tendency (mean, median), dispersion (standard deviation, range), and distribution (histograms, box plots).

Correlation Analysis: Correlation analysis will aid in determining the direction and strength of relationships between various variables, such as education, income, and occupation prestige; the Pearson correlation coefficient or Spearman rank correlation coefficient will be computed to quantify the associations.

Hypothesis Testing: Hypothesis testing methods, like t-tests or ANOVA, will be utilized to ascertain whether there are significant differences in occupation prestige scores among different demographic groups (e.g., gender, education level).

Regression Analysis: To model the link between predictor variables (such as education and income) and the outcome variable (occupation prestige score), regression analysis, including linear regression and logistic regression, will be employed. This will assist in determining the elements that significantly affect the prestige of a given occupation.  
  
  
Machine Learning Classification: Based on their attributes, machine learning classification methods, including support vector machines, random forests, and decision trees, will be used to categorize jobs into various prestige groups. This kind of supervised learning will shed light on the traits of high-status professions.

Cluster Analysis: Based on similarities in attributes (e.g., education, wealth), group jobs will be subjected to cluster analysis techniques such as k-means clustering or hierarchical clustering. This method of unsupervised learning will aid in the identification of discrete occupational groups with comparable prestige profiles.  
  
Text mining and Natural Language Processing (NLP): Textual data related to vocations, including job descriptions or census reports, will be analyzed using text mining and NLP techniques. Sentiment analysis, keyword extraction, and topic modeling are some of the tasks that will be included in order to obtain an understanding of the elements driving occupation prestige.

Data Visualization: To visually explore and communicate the analysis's conclusions, data visualization tools like as charts, graphs, and interactive dashboards will be used. We'll use visualization tools like Matplotlib, Seaborn, ggplot2, and Tableau to make visually appealing representations that facilitate comprehension and interpretation.  
  
  
By using these methods, the Ministry of Industry and Commerce in Sri Lanka will be able to make well-informed decisions regarding national development planning and obtain a thorough grasp of the elements influencing occupation prestige.

**Methodologies**

Methodologies for Occupation Prestige Analysis:

Exploratory data analysis, or EDA, is the process of examining a dataset in order to comprehend its organization, spot trends, and find anomalies. To obtain insights into the distribution and correlations between variables, dimensionality reduction (e.g., PCA), data visualization (e.g., histograms, scatter plots), and summary statistics will be used.  
  
Correlation Analysis: To determine the direction and intensity of correlations between occupation prestige and other factors like gender representation, income, and education, correlation analysis will be performed. To measure these associations, the Pearson correlation coefficient or Spearman rank correlation coefficient will be calculated.

Regression Analysis: To model the link between occupation prestige (the dependent variable) and predictor variables (such as income and education), regression analysis will be employed. Depending on the variables, either logistic regression or linear regression may be used. This analysis will assist in determining the key variables influencing occupation prestige.  
  
Employing cluster analysis, jobs will be grouped according to shared characteristics, such as gender representation, income, and educational attainment. To find different occupational clusters with comparable prestige profiles, methods like hierarchical or k-means clustering will be utilized. Understanding the variety of profession prestige across various industry sectors will be made easier with the help of this analysis.

Machine Learning Classification: Based on their attributes, machine learning classification methods, including support vector machines, random forests, and decision trees, will be used to categorize jobs into various prestige groups. With the use of supervised learning, it will be possible to discern patterns linked to the prestige of vocations and get insight into the traits of high-status careers.  
  
  
Analysis of Natural Language Processing (NLP): NLP methods will be applied to the examination of textual information related to professions, including job descriptions and census data. We'll carry out tasks like sentiment analysis, keyword extraction, and topic modeling to learn more about the variables affecting the prestige of a given occupation. This analysis will supplement quantitative data with additional context.

Statistical Testing: To ascertain whether there are statistically significant variations in occupation prestige scores among different demographic categories (e.g., gender, education level), statistical tests like t-tests or ANOVA may be utilized. Testing hypotheses will assist in confirming results and locating statistically significant correlations between variables.  
  
Data Visualization: To show the analysis's results in an understandable and visually appealing way, data visualization techniques will be used. Using programs like Matplotlib, Seaborn, ggplot2, and Tableau, charts, graphs, and interactive dashboards will be made to help in understanding and interpreting the findings.

The Ministry of Industry and Commerce will be able to make well-informed decisions on national development planning and obtain important insights into the elements that contribute to occupation prestige in Canada by utilizing these approaches.