# Analytics Startup Plan

**Synopsis: *This document provides a high-level walkthrough of the activities required to guide completion of the analysis.***

|  |  |
| --- | --- |
| **Project** | *Job descriptions* |
| **Requestor** | *Sara Memon* |
| **Date of Request** | *7/15/2024* |
| **Target Quarter for Delivery** | *8/15/2024* |
| **Epic Link(s)** | *https://www.kaggle.com/datasets/ravindrasinghrana/job-description-dataset* |
| **Business Impact** | 1. **Enhanced Recruitment Process**: Predicting salary ranges can help companies set competitive salaries and attract the right talent. 2. **Gender Diversity**: Understanding gender preferences in job postings can assist in addressing diversity and inclusion issues within the organization. 3. **Strategic Decision Making**: Data-driven insights can support strategic decisions in hiring, salary adjustments, and policymaking. 4. **Operational Efficiency**: Automated predictions can streamline the recruitment process, saving time and resources. |

## 1.0 Business Opportunity Brief

The analysis of the job description data set presents a significant opportunity to enhance the recruitment process through predictive analytics. By predicting salary ranges and gender preferences, companies can streamline their hiring strategies, promote diversity and inclusion, and make data-driven decisions. This can lead to more efficient operations, competitive salary offerings, and improved strategic planning, ultimately giving businesses a competitive edge in the job market

## 1.1 Supporting Insights

**Enhanced Recruitment Efficiency:** Automated predictions streamline candidate filtering, saving significant time.

**Competitive Salary Insights:** Predictive models ensure job offers are competitive, attracting top talent.

**Diversity and Inclusion:** Analyzing gender preferences helps identify and mitigate biases in hiring practices.

**Strategic Decision-Making:** Data-driven insights support informed and effective HR policy development.

**Operational Efficiency:** Automated processes reduce errors and enhance the scalability of recruitment efforts.

## 1.2 Project Gains

**Improved Hiring Decisions:** Companies can make more informed decisions based on predicted salary ranges and gender preferences.

**Increased Efficiency:** Automation of predictive tasks reduces manual workload and speeds up the recruitment process.

**Better Talent Attraction:** Competitive salary insights help attract and retain top talent.

**Enhanced Diversity:** Addressing gender preferences promotes diversity and inclusion within the organization.

**Strategic Advantage:** Data-driven insights provide a strategic advantage in the competitive job market.

## 2.0 Analytics Objective

**Predict Salary Range:** Develop a robust model to predict salary ranges based on various job-related attributes.

**Predict Gender Preference:** Create an accurate classification model to predict gender preferences for job postings based on job descriptions and related factors.

## 2.1 Other related questions and Assumptions:

**Data Completeness:** The dataset is complete and contains no critical missing values.

**Representative Sample:** The dataset is representative of the broader job market and includes diverse job postings.

**Static Market Conditions:** Market conditions remain relatively stable, making historical data relevant for future predictions.

**No Bias in Data:** The dataset does not contain significant biases that could skew the results of the predictive models.

## 2.2 Success measures/metrics

**Enhanced Recruitment Efficiency:** Predictive models can automate and streamline the recruitment process, saving time and resources.

**Competitive Salary Insights:** Predicting salary ranges helps companies offer competitive salaries, attracting top talent.

**Diversity and Inclusion:** Understanding gender preferences in job postings can help address and improve gender diversity within organizations.

**Strategic Decision-Making:** Data-driven insights support informed decisions in hiring, salary adjustments, and policymaking.

**Operational Efficiency:** Automation and predictive analytics reduce manual effort and improve the accuracy of recruitment processes.

## 2.3 Methodology and Approach

1. **Logistic Regression**: A basic yet powerful model for binary classification tasks, suitable for predicting gender preference.
2. **Decision Trees**: Useful for both regression and classification, providing interpretable models for salary prediction or gender preference.
3. **Random Forest**: An ensemble method that can improve prediction accuracy and handle large datasets with high dimensionality.
4. **Neural Networks**: Suitable for capturing complex patterns in data, though requiring more computational resources and careful tuning.

## 3.0 Population, Variable Selection, considerations

**Audience/population selection:**

**Observation window:**

**Inclusions:** Experience, Qualification, Salary Range, Location, Country, Work Type, Preference, Job Title, Role, Skills

**Exclusions:** Job ID,Latitude, Longitude, Company Size, Job Posting Date, Contact Person, Contact, Job Portal, Job Description, Benefits, Responsibilities, Company, Company Profile

**Data Sources:** Kaaggle.com

**Audience Level:**

**Variable Selection:** Experience, Qualification, Salary Range, Location, Country, Work Type, Preference, Job Title, Role, Skills

**Derived Variables:** None

**Assumptions and data limitations:**

## 4.0 Dependencies and Risks

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Likelihood (based on historical data)** | **Delay (based on historical data)** | **Impact** |
| **1: Data Quality**:  **2: Bias in Data**:  **3:** **Privacy Concerns**:  **4: Dynamic Market Conditions** | *Low*  *Moderate*  *High*  *Moderate* |  | Incomplete or incorrect data entries can lead to inaccurate predictions.  If the data set is not representative of the broader job market, the model may produce biased results.  Ensuring the privacy and confidentiality of contact details and personal information is critical.  Job market trends can change rapidly, which may affect the relevance of historical data. |

## 5.0 Deliverable Timelines

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item** | **Major Events / Milestones** | **Description** | **Scope** | **Days** | **Date** |
| 1. | Kick-off / Formal Request | *Initial meeting to formally start the project and understand requirements* |  | *1* |  |
| 2. | Assessment / Triage | *Evaluate the initial data and scope, identifying any immediate issues or roadblocks* |  | 3 |  |
| 3. | Prioritization | |  | | --- | |  |   Determine the priority of tasks and milestones based on project goals and timelines |  | 2 |  |
| 4. | Data Exploration & Analysis   * Issues with duplicates * Issues with Spend data | Initial exploration of data to identify duplicates and spend data issues. Identify and resolve duplicate entries in the dataset. Validate and correct any inconsistencies in spend data |  | 8 |  |
| 5. | Story Board 1 | Develop the first storyboard to outline the project's narrative and analytical approach |  | *2* |  |
| 6. | QA Output | Quality assurance review of the analysis output and storyboard. |  | *3* |  |
| 7. | Internal team Presentation | Present findings and storyboard to the internal team for feedback and adjustments |  | *2* |  |
| 8. | Go/No Go | Decision point to proceed with the project based on internal feedback and readiness |  | *1* |  |
| 9. | Story Board 2 | Refine and finalize the storyboard incorporating feedback from the internal presentation |  | 3 |  |
| 10. | Pilot | Conduct a pilot run of the project to test the findings and process |  | *7* |  |
| 11. | Delivery & sign-off | Final delivery of the project and formal sign-off from stakeholders |  | *2* |  |