**INTERVIEW QUESTION & SOLUTION**

**QUESTION 1:**

About the data & exercise:

* Columns 0 – 150 represent input variable and ‘target’ represents target variable
* Goal is to use the input variables to correctly identify or predict target variable
* Usage of Python and Jupyter notebook for completing this exercise is preferred but if you are not comfortable, feel free to provide source code and summary in email
* Please summarize your thought and analytics under following sections
  + Any pre-analytics steps and your understanding of data before you perform the analytics
  + Your choice of analytical algorithms, various steps taken during analytics, any comparisons between other algorithms and your understanding of how this model is performing
  + Your final choice of model and summary.

**ANSWER 1:**

* I have been given a dataset in the form of a csv file containing 150 input Variables & 1 Target Variable.
* Goal is to use the input variables to correctly identify or predict target variable.
* Detailed Summary of my Analytics needs to be Provided.
* Please find the Detailed Report below & My Report along with complete Source Codes & other Analytical Outputs have been uploaded in my GitHub Link and Google Drive Link mentioned below.
* **I HAVE CREATED A DETAILED REPORT AS BELOW CONTAINING ALL THE STEPS FOLLOWED , DETAILED OUTPUTS OF MY ANALYSIS & SUMMARY.**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **TITLE** | **PAGE NO.** |
|  |  |  |
| 1. | **OBJECTIVE OF THE PROJECT:............................................................................................**   * 1. **Problem Statement .................................................................................**   2. **Business Objective …………………………………………………………………………………**   3. **Technical Objective ……………………………………………………………………………….**   4. **Project Scope …………………………………………………………….............................**   **1.3 Need for A Solution: .................................................................................** | **3**  **4** |
| 2. | **COMPLETE STEPS INVOLED …………………………………………………………………………………………………………** | **8**  **9** |
| 3. | **BUSINESS UNDERSTANDING ……………………………………………………………………………………………** |  |
| 4. | * Project Design & Methodology |  |
| 3. | **DATA CLEANING & TRANSFORMATION PROCESS:**  **3.1 Data Type:** ........................................................................................................................  **3.2 Data Selection & Data Description:** ................................................................................  **3.3 Data Cleaning:** ..................................................................................................................  **3.4 Data Processing :** ..............................................................................................................  **3.5 Data Transformation :** .....................................................................................................  **3.6 Data Feature Engineering :** ............................................................................................... | **12**  **13 14**  **15**  **16** |
| 4. | **PROJECT DESIGN & METHODOLOGY.**  **(APPLYING BEST PRACTICES TO THE PROJECT OBJECTIVES & FRAMEWORK)**  **4.1 CRISP-DM framework along with SUCCESS ………………………………………………**  **4.2 Exploratory Data Analysis Project Design Architecture…………………….**  **4.3 Data Value Chain……………………………………………………………………………**  **4.4 Detailed Descriptive Analytics Methodology & Best Practices………………** | **17 18**  **19**  **20** |
| 5. | **GOVERNANCE : DATA GOVERNANCE & DATA MASKING & DATA SECURITY & PROTECTION**  **5.1 GOVERNANCE STEPS: Data Quality, Data Security & Protection, Data Stewardship**  **5.2 Tableau Server and Prep is used to perform Data Governance: Data Preparation, Quality Checking, Security, Stewardship and Data Quality …………………….**  **5.3 Tableau Governance ADD-ONS Data Management add on and Tableau Catalog Add on …………………………………………………………………………………………………………………** | **21 22** |
| 6. | **STATISTICAL ANALYSIS & TESTS AS A PART OF DESCRIPTIVE ANALYSIS ON DATA SET (EDA) USING PYTHON**  **6.1 Measure of central tendencies: Mean, Median, Measure of Spread, Skewness**  **& Kurtosis, Univariate and Bivariate Multivariate Analysis …………….** | **23 24** |
| 7. | **TABLEAU INTERACTIVE DASHBOARDS AND STORYBOARDS : DEPICT COMPLETE EDA ANALYSIS USING VISUALIZATION USING TABLEAU AND PYTHON………………………………………………………………** | **25 to 35** |
| 8. | **KEY FINDINGS AND BUSINESS INSIGHTS…………………………………………………………………………………….** | **36 37** |
| 9. | **KEY CONCLUSIONS………………………………………………………………………** | **38** |
| 10. | **RECOMMENDATIONS AND NEXT STEPS …………………………….** | **39** |
| 11. | **LIMITATIONS AND FURTHER SCOPE ……………………….** | **40** |
| 12. | **PROJECT TIMELINE : GANTT CHART .………………………….** | **41** |
| 13. | **CITATIONS & REFERENCES ……………………….** | **42** |
| 14. | **APPENDIX ………………………………………** | **43** |

**2.COMPLETE STEPS INVOLVED:**

* Business Understanding
* Project Design & Methodology
* Data Understanding
* Data Cleaning, Preparation & Transformation
  + - * Handling Missing Values
      * Handling Null Values
      * De-Duplication
      * Merging Datasets
      * Handling Outliers
      * Data Type Conversion
      * Data Quality Analysis
      * Normalization
      * Binning (or) Discretization
      * Handling Multi-Collinearity
      * Feature Engineering
      * Feature Selection (or) Dimension Reduction
      * Loading Data
* Initial Statistical Analysis & Inferences
  + - * Basic Statistical Analysis to Obtain Insights about Data
* Initial Descriptive/ Exploratory Analysis & Inferences
  + - * Graphical Analysis to Obtain Insights about Data
      * Correlation Matrix to Obtain Insights about Data
      * Association Analysis to Obtain Insights about Data
      * Cluster Analysis to Obtain Insights about Data
      * Segmentation Analysis to Obtain Insights about Data
* Predictive Modelling
  + - * Choosing Suitable Modelling Algorithms for Dataset & Justification
      * Following Pre-Data Modelling Requirements & Steps
      * Splitting Data into Train & Test
      * Steps Taken During Model Building using Each Algorithm
      * Model Validation/Testing & Scoring
      * Comparison of Algorithms & Justification
      * Model Performances Report
      * Choosing the Final best Algorithm/Model & Justification
      * Model Re-validation/Testing or Scoring
      * Performance Optimization & Prescriptive Analytics for chosen model
      * Model Re-validation/Testing or Scoring
      * Final Model Deployment
* Report Containing the Model Output & Key Insights Obtained.
* Data Governance & Protection Applied on Report & Publish Report.
* Summary & Conclusion of Analytics & Predictive Model Building performed .