

Integrated Capstone Project

This Case Study has four check points defined in it.

Check Point Topics	Remarks	Max Marks
1.1 Data manipulation using Python (50 marks) 1.2 Analysis using SQL Queries (50 Marks)	Checkpoint 1	100
2.1 Visualization using Power-BI (50 marks) 2.2 Data Analysis using Big Data Tools (50 marks)	Check Point 2	100
3.1 Data Analysis + ML Model Training and Deployment on Cloud (100 Marks)	Checkpoint 3	100
4.1 Final Presentation and Viva (50 marks)	Check point 4	50

Domain:

Workforce Analytics (HR)

About:

Introducing TalentXcel Solutions Pvt Ltd: Transforming Talent Acquisition in the Digital Era.

TalentXcel Solutions Pvt Ltd, a visionary start-up founded in Bangalore in 2010, emerged from a group of forward-thinking HR practitioners driven by the ambition to revolutionize the way organizations approach Talent Acquisition. With an unwavering commitment to staying ahead in today's fast-paced digital landscape, TalentXcel Solutions is poised to lead the industry into a new era of talent optimization.

As pioneers in end-to-end Talent Acquisition Solutions, TalentXcel Solutions combines cutting-edge technology, deep industry knowledge, and a passion for innovation to usher in a transformative shift in talent management practices. By leveraging advanced analytics, the company not only advises clients on their current Talent Acquisition landscape but also empowers them with individually tailored, data-driven strategies to propel their talent acquisition efforts to new heights.

Central to TalentXcel Solutions' success is their expertise in business process re-engineering, built on the pillars of People Capability, Process Maturity, and Technology Adoption. By thoroughly understanding the intricacies of talent acquisition in corporate environments, the company seamlessly integrates these three tenets to deliver customized, enterprise-class solutions that effectively address the unique talent challenges faced by their clients.

With TalentXcel Solutions as a strategic partner, organizations gain access to a comprehensive suite of services designed to optimize their talent acquisition processes. From talent sourcing and candidate screening to performance evaluation and retention strategies, TalentXcel Solutions brings together the best practices and innovative tools required to attract, engage, and retain top talent in the digital age.

Embracing the ethos of continuous improvement, TalentXcel Solutions stays at the forefront of emerging trends and technologies in Talent Acquisition. By providing a holistic approach that combines cutting-edge analytics, strategic recommendations, and seamless implementation, TalentXcel Solutions enables its clients to thrive in the ever-evolving talent landscape and achieve long-term success.

In summary, TalentXcel Solutions Pvt Ltd is a dynamic and forward-looking start-up that harnesses the power of technology, analytics, and business process re-engineering to deliver transformative Talent Acquisition Solutions. By embracing innovation and taking a holistic approach, TalentXcel Solutions empowers organizations to navigate the complexities of talent management in the digital era and unlock their full potential.

Challenges:

TalentXcel Solutions Pvt Ltd understands that client service heavily relies on the quality of the people involved in delivering business outcomes. However, one of the prominent challenges faced by HRWorks and its clients revolves around effectively managing a high-quality workforce. Despite investing significant time and effort into creating a supportive and thriving work environment, organizations often experience employee attrition as individuals seek better opportunities elsewhere. This cycle of attrition necessitates constant recruitment efforts, resulting in a continuous loop of hiring and the associated costs and disruptions.

To address this challenge, HRWorks and its clients have recognized the need to find a unique solution that goes beyond the conventional process-oriented aspects of human resource management. Understanding the significance of identifying potential attrition cases in advance, HRWorks has put forth a subtle yet crucial proposal: "A significant proportion of candidates decline job offers after acceptance. If we can accurately identify these candidates in advance, companies can proactively allocate their resources more efficiently."

TalentXcel Solutions Pvt Ltd acknowledges the importance of bridging the gap between attrition and recruitment. In response to this challenge, TalentXcel Solutions aims to offer a comprehensive solution that extends beyond standard recruitment processes. By leveraging data analysis and predictive modeling techniques, TalentXcel Solutions aims to develop a robust predictive model that can accurately assess the likelihood of a candidate joining a company after accepting a job offer. This unique approach will enable organizations to optimize their talent acquisition strategies, minimize resource wastage, and create a culture of inclusivity and stability.

By partnering with HRWorks and their clients, TalentXcel Solutions Pvt Ltd aims to provide a valuable solution that will significantly improve workforce management and help organizations break free from the cycle of attrition and continuous recruitment.

What is Expected?

As a data analyst at TalentXcel Solutions Pvt Ltd, I have conducted an exploratory analysis on the data provided by HRWorks to understand the challenges they face in predicting the likelihood of a candidate joining the company. The dataset consists of information about the candidates who accepted job offers, including candidate demographics, education, experience, interview ratings, and other relevant variables. During the analysis, I identified potential issues such as missing values, outliers, and data imbalance that need to be addressed. In terms of descriptive analysis, I examined candidate demographics, education, experience, interview ratings, and time-to-join to gain initial insights.

Based on these findings, the next steps involve cleaning the data, performing feature engineering and selection, building, and evaluating predictive models using different

approaches, fine-tuning the selected models, interpreting the model's feature importance's, and ultimately deploying the model for predictions. By following this comprehensive approach, TalentXcel Solutions Pvt Ltd will be able to develop an effective predictive model to classify candidates as "joined" or "not joined" and gain valuable insights into the factors influencing candidate decisions.

Data Dictionary:

- **Sno:** SI number auto increment
- **Candidate_Ref:** Candidate reference number
- **DOJ_Extended:** Date of Joining of extended
- **Duration_to_accept offer:** Duration to accept the offer by candidate
- **Notice_Period:** Notice period of previous employer
- **Offered_Band:** E1 < E2 < E3 and so on
- **Percent_hike_expected_in_CTC:** expected hike by candidate
- **Percent_hike_offered_in_CTC:** hike offered by joining organisation
- **Percent_difference_CTC:** difference between expected and offered
- **Joining_Bonus:** any joining bonus offered.
- **Candidate_relocate_actual:** relocating required or not
- **Gender:** Gender of candidate
- **Candidate_Source:** How candidate applied or reached
- **Rex_in_Yrs:** years of exp
- **Location_ID:** Code for current location of organisation
- **Postal_Code:** Postal Code of the area of organisation
- **Domicile_Id:** Code for home state of the candidate
- **Region_Code:** Code representing unique Region name
- **Age:** Age of candidate
- **Date_of_Interview:** Date on which Interview was conducted.
- **LOB_id:** Unique id for Line of Business (LOB)
- **LOB:** Line of business

Target variable:

- Status: joined or not.

Check Point 1

Task 1.1(Data Manipulation using Python)

Here are some indicative types of analysis you can perform. Please note that this is not an exhaustive list, you may add more

- Come up with appropriate results for the following:
 - Analysis of percentage joined of offer released.
 - What are the key drivers that influence the candidate joining/not joining a company?
 - Are there specific locations where candidates are not joining?
 - Joining status depends on the duration to accept an offer?
 - Hike offered has an impact on joining status.?

Task 1.2 (SQL-Oracle)

Stage 1:

- Construct an ER-Diagram for the above-mentioned Requirement
- Construct Tables as per the ER-Diagram.

- Identify the relationships between tables and use appropriate standards for the same where applicable
- Insert the appropriate data into the identified tables from the sample dataset provided.

Stage 2:

- Generate Info those candidates who have accepted offer and joining time is less than 30 days and candidates who are ready to re-locate.
- Generate Info those candidates who have accepted offer and display list the candidates who have been offered and yet to accept the offer within 10 days.
- Generate Info those candidates who are willing to join and the ECTC is 25% hike from their CTC.
- Generate Info those candidates who are willing to join and the ECTC is 25% hike from their CTC joining time is less than 30 days and candidates who are ready to re-locate and joining bonus is offered.
- Generate the count of the candidates who are hired through what source and also who have joined and declined the offer.

Deliverables/Submission guidelines of Checkpoint 1

1. You must prepare a power point presentation with screenshots of outputs (10 -15 slides) for each check point
 2. Mention Problem Statement and Your approach to the problems
 3. You need to submit all the code files - Task 1.1
 4. The code file (html file for Task 1.1) should contain the Batch Name and the group name, group members (One of the group members) at the top (in Jupyter Notebook).
 5. All comments/inferences/insights/reasons for doing a particular tasks etc should be written as a 'markdown text', but **NOT** using a comment line with # or ''.
 6. Submit the code file as HTML file format (you have an option in Jupyter Notebook to save the file as HTML).
- Name of the file must be in the form of:
BatchName_FirstName_SecondName.html
7. Task 1.2 SQL code to be copied in the word doc
 8. The presentation file should have the Batch name, group name, Project name, Group members, their responsibilities
 9. Upload all the deliverables in the UNext LMS

Check point 2 (Visualization using Power-BI, Data Analysis using Big Data Tools)

TASK 2.1(Visualization using Power-BI)

Connect the data with Power BI desktop and perform Data Manipulation using Power Query Editor. Perform the below tasks in Power BI Desktop.

- Which gender is having the highest number of experiences in the dataset?
- Which location has witnessed the highest number of joiners? How are the joiners compared across different locations?
- Identify the Gender taking the longest time to accept the offer from the company.
- Indicate the Gender earning the highest Average of percent hike in CTC.

NOTE: Results and graphs must be backed with appropriate inferences and insights.

TASK 2.2 Data Analysis using Big Data Tools

What is Expected?

Big Data technologies like HDFS, Hive and PySpark need to be used as the historical data increases in size. As part of this task the following activities need to be done.

- Develop a PySpark application to load data Spark DataFrames and save it into Hive tables on a Hadoop cluster in an optimized format.
- Write PySpark routines to cleanse the data, prepare the data to handle missing values, and the data transformations identified in task 1.1 again making sure that the data is written into Hive tables in an efficient format
- Ensure that the best practices are followed, and the design & code use the features of Spark and take advantage thereof.

Deliverables/Submission guidelines of Checkpoint 2

1. You must prepare a power point presentation with screenshots of outputs (10 -15 slides) for each check point.
2. Mention Problem Statement and your approach to the problems
3. Task 2.1
 1. PowerBI .pbix file to be submitted.
 2. Have all comments written properly in the .pbix file.
 3. The .pbix file should contain the Batch Name and the Group Number, Group member names at the top.

Task 2.2

- Submit Jupyter code file in html format. The code file (html file for Task 2.2) should contain the Batch Name and the group name, group members (One of the group members) at the top (in Jupyter Notebook).
 1. All comments/inferences/insights/reasons for doing a particular tasks etc should be written as a 'markdown text', but **NOT** using a comment line with # or ''.
 2. Submit the code file as HTML file format (you have an option in Jupyter Notebook to save the file as HTML).
 3. Name of the file must be in the form of:
BatchName__GroupNumber_FirstName_SecondName.html
 - 4.
- Put all Tasks 2.1 & 2.2 as zip file (Mentioning batch name, Group number and your name) and upload it on the LMS.

Checkpoint 3

Task 3.1 - Data Analysis + ML Model Training and Deployment on Cloud

AZURE

1. Azure Synapse to PowerBI Connectivity
 - a. Move the DataSet to Azure Synapse Storage Gen2
 - b. Create a serverless SQL pool to query the data from Storage gen2-autogenerated script
 - c. Create a Linked service to PowerBI
 - d. Ensure you have sufficient privileges on Synapse to access the serverless SQL pool.

- e. Perform various analytics on PowerBI – at least 3 dashboards
2. Enable Azure blob storage monitoring by adding sample data and upon processing if storage receives more than 20 bytes of data.
3. Azure blob to azure data bricks using notebook options databricks to powerbi connectivity.
4. Azure blob to Azure SQL copy option using datafactory and connect Azure SQL to Databricks.
5. Write Azure functions to trigger when blob storage exceeds 20 bytes of data.
6. Build appropriate ML model/s on the data using Azure Machine Learning, Identify the right metric to evaluate the performance of the model **and Deploy on Azure Machine Learning**

Deliverables/Submission guidelines of Checkpoint 3

Task 3.1

Complete all the above tasks on your respective Cloud Platform allotted and for submission take screenshots of each task specified with step-by-step flow in a word document with proper caption mentioned along with your Batch/Group/Team member names convert as a PDF file and submit the PDF document on the LMS

Checkpoint 4

Task 4

Prepare crisp Final presentation including all three Checkpoint achievements and appear for Q&A session

Deliverables/Submission guidelines of Checkpoint 4

- You must prepare a power point presentation with screenshots of outputs (10 -15 slides)
- Submit the ppt.