SQL Project: Competitive Landscape

Overview: Ironhack has hired us to conduct an analytics consulting project titled "competitive landscape". We created and populated a comprehensive database with coding schools that were direct competitors to Ironhack, with focus on the top 3 Data Analytics Bootcamps.

This project is centered around a meticulous analysis of Ironhack's competitive position within the coding schools. Our primary initiative involved the establishment and population of a comprehensive database containing coding schools identified as direct competitors to Ironhack.

Business Question:

How can the performance reviews for Ironhack students be improved?

Getting Started:

* Data Collection:

Our first step was to collect data not only from Ironhack, but also from other schools, ensuring a comprehensive grasp of the current performance review process and identify areas for improvement.

* Entity-Relationship Diagram (ERD):

In our case, after we collected the data, we preferred to have Entity-Relationship Diagram to clearly define the relationships and connections between the different entities within the data, and to facilitate the organization and management of the data in a structured and efficient manner.

* DATA CLEANING:
  + Rigorous data cleaning was required in order to ensure its accuracy and consistency, and to eliminate any discrepancies or irrelevant details that may affect the analysis and interpretation of the results.
  + During the data cleaning phase, more efforts were made to align all columns with their counterparts in the SQL database schema.
  + To enhance data quality, Comment dataframe underwent a process of row deletion to eliminate instances with null values
  + In some dataframes, an 'id' column was introduced for specific tables, leveraging the dataFrame index and mapping techniques. This strategic addition enhances data organization and facilitates efficient indexing for subsequent database operation.
* Analysis & Sentimental Analysis:

Detailed analysis of the collected and cleaned data involves deploying diverse statistical and data analysis techniques. Additionally, conducting sentimental analysis evaluates the impact of varied assumptions and scenarios on the derived results and conclusions.

* Derive Answers:

This process involves extracting meaningful insights from the data analysis results by identifying patterns, discerning trends, and relating them to the original business questions and objectives pertaining to the performance review process form the core of interpreting the findings.

**Our Recommendations:**

* + - Regularly review and update the studying program to align with industry trends and technologies.
    - Use feedback of Alumni and professionals to be sure that program is relevant.
    - Introduce students to a broad spectrum of technologies and tools that are currently in demand in the IT job market.
    - Improve job support (invite guests from IT job positions, facilitate networking opportunities with potential employers and professionals, mentor program)

Techniques Used:

* Web Scrapping
* Data Cleaning
* Entity-Relationship Diagram(ERD)
* SQL(Structured Query Language)
* Analysis, including sentimental analysis

Project Deliverables:

The deliverables for this project are:

* .ipnby file solution submitted via github
* An exported .sql file with schema
* Readme file
* Presentation