

The Data Challenge has two different areas of analytics skill assessment. Please see the Excel file attached along with this document.

- 1) Descriptive Analytics**
- 2) Predictive Analytics**

Submit your responses in a word document(SQL Queries, Summary of findings) and scripts (Python/R) with references to the Question numbers.

Assumptions

1. The tab names in the Excel file are the names of the tables in SQL database and the column names in SQL tables are exactly same as column names in the excel file.
2. Assume The tabs “Market_Aug2021” and “Market_Aug2018” represent the complete market as of Aug2021 and Aug2018 respectively.
3. If anything is not stated, Please make appropriate assumptions as required and state the assumptions in your response.

Descriptive Analytics

Data set Description

Refer to tabs “Market_Aug2021” and “Market_Aug2018” in the attached excel file. Convergence has various data sets of different Investment funds (Identified by unique reference “FundID” and has a Dollar amount in “Assets”) in the market that are managed by different investment banking companies (Also referred as “Adviser” in the data set, has a unique identity “CRD”). The Advisers of these investment funds use different service providers and pay them a service fee to manage the funds operationally. In this instance, the data set provides one such service provider, the Auditing company (Referred as “Auditor”) which is appointed to perform the financial audit for each of the fund. Like any business, the auditing companies prime objective is to audit as many funds as possible, be a competitive service provider thereby increase their market share and revenue from the audit business.

Questions

1. Write a SQL query (or sequence of queries) to determine the market share(%) of each Auditing company as of Aug2018 and Aug2021. Market Share is determined in two different ways and output both metrics in two different columns in the result. Result should include columns Auditor, Marketshare_Funds, MarketShare_Assets.
 - a. MarketShare_Funds (%) by Number of Funds the Auditor Audits
 - b. MarketShare_Assets (%) by the Aggregate \$ amount of Assets that they audit

2. Using the above result, write a SQL query (or sequence of queries) to determine the Top 3 biggest gainers and Top 3 biggest losers in change in market share in Funds from Aug2018 to Aug2021.
3. You will notice in the data set, that there are funds that are newly disclosed in “Aug2021” and that did not exist in “Aug2018” and funds that are no longer disclosed from “Aug2018” in “Aug2021”. The result for each of the following questions should include CRD, Adviser, FundID, FundName, Assets, Auditor
 - a. Write a SQL query to determine New Funds that are disclosed in “Aug2021” but did not exist in “Aug2018”
 - b. Write a SQL query to determine Dropped Funds that are disclosed in “Aug2018” but are dropped in “Aug2021”
 - c. Write a SQL query to determine Change in Assets (Difference in asset value from 2018 to 2021) for funds that are disclosed both in “Aug2018” and in “Aug2021”. This is also referred as “Asset Inflation/Deflation”. The result should include “ChangeinAssets” column instead of “Assets” column.
 - d. Write a SQL query to determine funds that changed an Auditor between Aug2018 to Aug2021. Indicate who is the Auditor that took the most business from its competitors and Indicate the auditor that lost the most business to its competitors.
4. Adviser is a “client” of the Auditing company disclosed in the data set. An Adviser may advise one or more funds.
 - a. Write an SQL query to determine the number of unique clients each Auditor has in Aug2021. The result set should include columns “Auditor” and “NumberofUniqueClients”.
 - b. Extend the above 4(a) SQL query to only display the Auditors that have atleast 5 unique clients.
5. The larger the asset size of a fund the Auditor provides a service, the more revenue the Auditing company makes as they get paid on percentage of fund assets Audited. Write a SQL query to create a calculated column “AssetSizeBand” in Aug2021 that has coded values “0-\$1Bn”, “>\$1Bn-2Bn”, “>\$2Bn-3Bn” and “>3Bn” based on the “Assets” column. The result should include columns FundID, FundName, Assets, AssetSizeBand
6. Write a SQL query to filter “Market_Aug2021” table to output only Funds that have Assets greater than \$3Bn. The result can include all columns of the table. Also the result should be sorted with funds from largest to smallest assets.

Predictive Analytics

Data set Description

Refer to tab “LikelytoChurn” in the attached excel file. Like any other industry, A common phenomenon in Fund services business is that the Advisors change their service providers for better service, which is synonymous to classic customer churn data science problem.

The data set consists of all Advisors (Identified by unique identity “ID”) that changed a service provider and all Advisors that did not change a service provider indicated by column “Switched” as “Yes” or “No” respectively. There are various factors that describe the nature of the advisors business and are indicated as “Factor#1” to #Factor#38”. The purpose of Predictive analytics exercise has two objectives.

- 1) Determine which Factors influence the outcome of changing (“switching”) a service provider the most and describe why using various statistical exploratory analysis techniques. In the interest of time, You can limit your description to Top3 factors that influence the most.
- 2) Build a Classification predictive model to predict the likelihood of an Advisor changing (“Switching”) a service provider. You could choose to split the data into 70% Train and 30% Test data sets with the aim of training the model with “Train” data set and testing the model on the “Test” data set. The final result should include “Confusion matrix” on the “Test” data set and any model validation metrics that you chose to include. Write a report on your approach and the summary of your observations. The aim of this exercise is to convince a business user of the predictive value and take an appropriate action to avoid losing business from Advisors.