DAT 494 – Final Lab Assignment (200 points)

Database : Latham database

Objective: Use Excel spreadsheets from the Latham database tables to create Alteryx workflows and Tableau visualizations in order to analyze baseball data.

Data: Choose a MLB team from the Google sheet. (Note: Do not include the 1981, 1994, 1995, and 2020 seasons as they were shortened seasons with less than 162 games)

Alteryx Workflows (100 points) -

> **For your team only,**

1. (30 points) Teams Analysis
   a. In the same workflow (using multiple branches), using the *TEAMS* table:
      i. (10 points) Find the average number of wins (W) by each team in your team's league (AL or NL) between 2000-2019. Create a report and highlight your team. Where does your team rank?
      ii. (10 points) How successful is your team? Find the number of times your team has achieved winning the Wild Card (WCWin), Division (DivWin), League (LgWin), and World Series (WSWin). Create a report.
      iii. (10 points) Using the Rank field, create a report that identifies how many times your team has ranked between 1st and 5th place in their division between 2000-2019.
2. (35 points) Salaries Analysis
   a. (5 points) Using the *Salaries* spreadsheet include the records for your team (Note: the time period for this table is between 1985 and 2016), find the minimum and maximum salary for each year.
   b. (5 points) Create a new field "Key Player" using a conditional statement in a Formula that should be set to "Min" for the player(s) with the minimum salary, "Max" for the player(s) with the maximum salary and "Neither" for all other players.
   c. (5 points) Join the PEOPLE spreadsheet to include the player's name.
   d. (5 points) Include only the players that represent the minimum and maximum salaries.
   e. (5 points) Sort the information by year.
   f. (10 points) Create a report that includes: Team Name, Year, Minimum Salary, Maximum Salary, Player Salary, Player Name.
3. (30 points) Attendance Analysis
   a. (5 points) Using the HomeGames spreadsheet, find the average attendance for each ballpark for your team between 1969-2021 (excluding 1981, 1994, 1995 and 2020).
   b. (5 points) Calculate the difference between each season's attendance and the average attendance for the ballpark.
   c. (5 points) Remove any ballparks with less than 5 games.
   d. (5 points) Join the PARKS spreadsheet to include the ballpark name.
   e. (10 points) Create a report with appropriate information and export the report to Excel.

Submission Requirements – (5 points) Include all workflows for each step and related output. Upload the screenshots of the Alteryx workflows and results window for each question into **one PDF**. Additionally, upload **one Excel spreadsheet** with different sheets that represent the results for each part of the process.

Tableau Visualizations (100 points) -

For the parts 1-4 use the 2000-2019 seasons only,

1. (20 points) Using the Teams spreadsheet, create a line graph showing the number of wins for your team. Add total number of runs as a dual axis graph using a different type of graph. Add a reference line to indicate the average number of wins over this time period.
2. (15 points) Using the Teams spreadsheet, create a visualization that displays your team's season BA over this time period. You will need to create a calculated field for BA, which can be calculated as H/AB. Adjust the format of BA to 3 decimal places. Adjust the y axis to make the visualization better.
3. (15 points) Using the Teams spreadsheet, create a visualization that compares your team's number of wins with the other teams in their division (DivID) over this time period. You may also need to create groups to combine teams that changed names. You should have a maximum number of 5 teams on your graph. Use a distinct color to represent your team.
4. (15 points) Using the Salaries spreadsheet, create a visualization to compare the average annual salary for each season comparing the NL vs. AL (LgID).
5. (15 points) Using the Teams spreadsheet, for the **1960-1969 seasons**, create a visualization (Note: this could be a table) to show the total wins for each team. Sort the graph in descending order to identify the team with the most wins in this decade. Use the team name in the visualization and color to enhance the graph.
6. (10 points) Build a dashboard that includes at least two of your visualizations.

Guidelines for your visualizations:

- Title should be appropriate for the graph.
- Axis names should be changed where necessary.
- The graph should be visually appealing, so be careful with the use of color.

Submission Requirements – (10 points) For each problem, you need to include a screenshot of the Tableau sheet with the row and column heading and any filters used clearly visible. These should be loaded into **one PDF**. Each visualization should be copied into a **Powerpoint presentation** and uploaded to Canvas.

Submission Guidelines – Submit the required output (2 PDFs, 1 Excel spreadsheet, 1 Powerpoint presentation) to Canvas by Sunday, December 3 at 11:59 PM. No late submissions will be accepted.