LA Clippers Data Analytics Intern Data Challenge

In the 2018/2019 NBA Season, we played 82 games. Of those 82 games, 41 of those games were home games. We are responsible for determining pricing for those 41 home games and part of the process is to consider attendance, which can be influenced by a variety of factors. An example of a typical project as a BI Intern would be to clean and analyze historical data at our home games to help us determine how we price our games. For this interview, you have been provided sample data to (1) answer some of the questions below and (2) develop a model to forecast attendance as if you were forecasting before the season had started. You are expected to present your model and findings in PowerPoint. The following are more detailed descriptions of the data provided, what we want you to turn-in, and how we are going to assess your performance/candidacy with regards to the Data Challenge.

Data:

1. Ticket scans – Actual attendance to our games.
2. Pricing data – How we have priced our seats/games.
3. Seating chart data – Where each section is located.
4. Game data – Details regarding all the 2018/2019 games played and the outcome.
5. Team data – Other data for each team

For deliverables, please provide the following:

- **Part 1: Word Document** answering the following questions with graphs or tables:

  1) **[SQL]** Do weekend games (Friday, Saturday, Sunday), on average, have a higher attendance than weekday games (Monday through Thursday)?

  2) **[SQL]** Identify and rank the top 4 opponents with the highest average number of attendances.

  3) Identify and rank the top 10 sections that are, on average, the most filled to their capacity.

  4) Is there a correlation between the opponent team having a higher Vegas Odds Score (indicating higher probability of winning a championship) and higher attendance?

  5) On average, which 15-minute period before or after the start of a game has the highest number of people scanning. (Example 30 – 16 minutes prior to the start of a game)

    o PLEASE NOTE: scan date and time is in Central Time Zone; event date and time is in Pacific Time Zone.

Note: For [SQL] questions, please follow this example to use the SQL query in Python:

!pip install pandasql
from pandasql import sqldf
import pandas as pd

team = pd.read_excel('Interview Project Data.xlsx', sheet_name='Team Data')
game = pd.read_excel('Interview Project Data.xlsx', sheet_name='Game Data')
seat = pd.read_excel('Interview Project Data.xlsx', sheet_name='Seating Chart')
price = pd.read_excel('Interview Project Data.xlsx', sheet_name='Ticket Price Data')
scan = pd.read_excel('Interview Project Data.xlsx', sheet_name='Ticket Scan Data')

example = sqldf("select g.*, t.'Team Code' from game g left join team t on g.Opponent = t.'Team Full Name'")

Please see the screenshot as an example:

- **Part 2: PowerPoint Presentation with Video/Screen Recording**:

    1) Deliver 1~2 insights that you think are interesting about the dataset.

        o   Be creative and visualization always helps!

    2) Develop a model to predict the at regular season games attendance.

- o Imagine it is 2 weeks before the 2019/20 season tip off date and you are given the task to predict how full the stadium will be for the coming season.
- o If you plan to incorporate outside data, please list the data sources and explain how you plan to collect the data.
- o While accuracy is important, this is not a Kaggle contest. Please see below for more details on what we are looking for.
- o There is no limit to the number of slides, but please be succinct. You are free to have as many exhibit/appendix slides as you'd like.
- o Be sure to explain:
  - ▪ How, what, and why you chose your approach
  - ▪ What features did you use and why?
  - ▪ What are your significant findings?

3) Make a Video/Screen Recording to present your findings.

- o Be clear and concise (5 minutes or less).
- o Focus on how you would present your model and results to non-data people.
- o Make sure to showcase your non-technical skills in this portion (communication ability, design, creativity, etc.).

- **Part 3: A copy of your work sheet** where you performed your analysis (Excel workbook, Jupyter notebook, etc.)

For your reference, we are looking at the following when assessing your candidacy:
- Technical Skills
  - o Solving the problem with the tool of your choice (preferably, Python)
  - o Clarity and efficiency as shown in your work sheet provided
- Analytical Thinking
  - o What approach (metrics and models) you use and why
  - o Accuracy, though not as important as approach
- Communication skills
  - o How well can you explain your findings at a high level and defend them on a technical level
- Design Skills
  - o PowerPoint presentation is designed with clarity and purpose.
- Creativity
  - o Inclusion of outside data sources and information
  - o How deep you look at the data given.