**SUMMARY**

Major League Baseball hasn’t had a league expansion since 1998 and I have always been curious to know which 2 city’s would be best suited for hosting when the league expanded to 32 teams. This analysis seeks to examine the top 10 cities in the US and based on population, specific age ranges, income levels, and ethnic breakdowns, and see how they compare to one another.

Having grown up playing little league, I followed MLB into my adulthood and look forward to when the league decides to grow into other cities. Throughout the project I used Excel, Python and PowerBi.

**QUESTION**

This project will analyst the 10 cities compared to average statistics from the 30 cities that currently have MLB teams. By utilizing census data from cities with MLB teams, I will discover common characteristics across several categories and use them as benchmarks to see how the 10 cities that don’t have teams compare. Hopefully it will help in predicting where the league will attempt to put the next 2 teams when it expands.

## DATA SOURCES

1. Census Reporter [Census Reporter: Making Census Data Easy to Use](https://censusreporter.org/)
2. Government Census [Census.gov](https://www.census.gov/)
3. Baseball Reference [MLB Stats, Scores, History, & Records | Baseball-Reference.com](https://www.baseball-reference.com/)
4. Statista [MLB fans in the U.S. by ethnicity 2020 | Statista](https://www.statista.com/statistics/1147621/mlb-fans-ethnicity/)
5. SponsorPulse [MLB Fan Demographics: Who’s engaging with the MLB in America | SponsorPulse](https://www.sponsorpulse.com/insights/mlb-fan-demographics-whos-engaging-with-the-mlb-in-america)
6. ESPN [All 30 MLB Teams: Location, Stadium and Website Information | MLB.com](https://www.mlb.com/team)

**PROBLEMS AND HURDLES**

I originally wanted to include Mexico and Canada when considering various cities but due to their census data being dated differently as well as having been divided into broken down differently, it wasn’t possible to compare equally.

Also, there are other factors that I couldn’t take into account that help determine locations for new teams that census data cant provide like ownership groups having a say as well as cities giving approval.

**TECHNOLOGIES USED**

1. Excel – Initially used Excel to order and narrow down the data that I needed and got rid of the information that was excess.
2. Python – Utilized Python to make charts to visually represent the data, specifically within a Jupyter Notebook.
3. PowerBI - Upon completion of normalization and analysis, PowerBI was utilized to create visualizations and an interactive dashboard map.

**CONCLUSION**

Austin Texas and Charlotte North Carolina make the two best places for MLB to put expansion teams (based on the analyzed data).