**Project Title: Evaluating Optimal States for Health Tracker Smartwatch Business Expansion**

Date of Submission: September 4, 2023

Name: JIGGY BOY G. ROMO

**Introduction**

In today's rapidly evolving business landscape, strategic expansion decisions are paramount to the growth and sustainability of a company. As our startup company endeavors to flourish, a key opportunity presents itself in the form of expanding our health tracker smartwatch business. The objective of this project is to meticulously identify and evaluate the state or states within the United States that offer the greatest potential for the successful expansion of our health tracker smartwatch line.

**Problem Statement**

As our startup company strives to expand and grow, a critical challenge arises in selecting the most opportune state or states for the expansion of our health tracker smartwatch business. The dynamic landscape of wearable technology and health-conscious consumers demands a well-informed decision that maximizes market potential and ensures sustainable growth. Therefore, the primary problem at hand is to identify and recommend the state(s) within the United States that offer the most promising environment for the successful expansion of our health tracker smartwatch line.

**Objectives**

The objective of this analysis is to leverage the available data on state income and corruption rate to identify and recommend the state(s) within the United States that hold the most favorable condition for the successful expansion of our health tracker smartwatch business. By conducting a comprehensive evaluation of these two key factors, we aim to make an informed and data-driven decision that aligns with our growth goals and ensures the viability and profitability of our expansion strategy by determining the following:

* Percentage of income by state and corruption conviction if there any observable connection.
* Identify the states with the highest and lowest average income.
* Identify the states with the highest and lowest corruption conviction rates.
* Test the correlation between state income and corruption convictions.

****

**Project Title: Evaluating Optimal States for Health Tracker Smartwatch**

**Business Expansion**

**Methodology**

1. **Data Extraction and Transformation:**

The initial step involves the retrieval and processing of data. Commencing with the restoration of the PostgreSQL database backup file, which contains economic indicators data, is essential. Following the successful database restoration, SQL queries are employed to extract and reshape the data from the database tables. Subsequently, the extracted data is transferred to an Excel format to facilitate further analysis.

1. **Data Cleaning and Preparation:**

* **To verify the presence of DUPLICATE rows in the `state\_income` table using the `average\_income`, `minimum\_income`, and `maximum\_income` columns, I employ the subsequent SQL query: I select each column, count the occurrences, group the data by each column individually, and finally filter for counts greater than one. In the end, the query reveals that the table contains no duplicate values.**
* **To verify the presence of NULL values in the table, I began by selecting all records from the `state\_income` table where either `average\_income`, `minimum\_income`, or `maximum\_income` is null. The outcome of this query revealed that there are no null values within the table.**

1. **Data Analysis:**
2. **SQL**

* Identify the states with highest and lowest average income. To pinpoint the states with the highest and lowest average incomes, I employ the following approach. I create two temporary tables: one for the states with the highest incomes and another for the lowest. These tables are constructed by selecting `state\_usa`, `average\_income`, and labeling them as 'Highest' and 'Lowest' accordingly. The results are sorted in descending order for the highest incomes and ascending order for the lowest, with a limit of five states each. Finally, I combine the results from both tables using the UNION ALL operator to present a comprehensive list of states with their respective income categories.
* Identify the states with the highest and lowest corruption conviction rates. To identify the states with the highest and lowest conviction rates related to corruption, I follow this approach. I create two temporary tables: one for states with the highest conviction rates and another for the lowest. These tables are formed by selecting `state\_usa`, `convictions\_per\_capita`, and labeling them as 'Highest' and 'Lowest' conviction rates accordingly. The results are sorted in descending order for the highest conviction rates and ascending order for the lowest, with a limit of five states in each category. Finally, I merge the results from both tables using the UNION ALL operator to present a comprehensive list of states along with their respective conviction rate categories.

1. **EXCEL**

* To visualize the findings in Excel, I create a scatter plot where I place the 'convictions per capita' column on the y-axis and 'average income' on the x-axis. I enhance the visualization by adding a trendline to the plot, which depicts a downward trajectory.
* For additional examination within Excel, I generated a bar chart to identify states with the highest average income, emphasizing the top 5 states. In this chart, I placed 'average income' on the y-axis and 'state USA' on the x-axis. The outcome revealed that the top

****

**Project Title: Evaluating Optimal States for Health Tracker Smartwatch**

**Business Expansion**

5 states with the highest average income are Maryland, Massachusetts, New Jersey, California, and Connecticut.

1. **PYTHON**

* In a Python data visualization context, I constructed a scatter plot to depict the correlation coefficient between state income and convictions per capita, yielding a value of -0.23. The process entailed calculating the correlation coefficient using the following syntax: `correlation\_coefficient = merged\_df['average\_income'].corr(merged\_df['convictions\_per\_capita'])`, followed by the creation of the scatter plot.
* I also generated a correlation matrix as part of the analysis. This was achieved by employing the following code: `correlation\_matrix = merged\_df[['average\_income', 'convictions\_per\_capita']].corr()`. The outcome revealed a correlation coefficient of -0.23. Following this, I proceeded to create a heatmap using Seaborn to visually represent the correlations.
* A correlation coefficient of -0.23 indicates a weak negative correlation between ‘average income’ and ‘convictions’.
* Finally, I crafted a histogram to illustrate the distribution of income across all states. This involved setting the x-axis label as 'Average Income' and the y-axis label as 'Frequency'. The histogram graph displayed a positive skewness, indicating that the majority of states possess lower average incomes while a minority of states exhibit considerably higher average incomes.

1. **POWER BI**

* In Power BI, I showcased the states with the highest and lowest convictions by presenting the top 5 states with the highest convictions and the top 5 states with the lowest convictions. Additionally, I created a visualization map with bubbles to represent the conviction values for each state.
* In the end, I generated a correlation matrix examining the relationship between income and convictions within each state. The outcome of this analysis indicates a mild negative correlation between income and conviction rates. In other words, as income levels increase within the dashboard, there is an observable trend where conviction rates tend to decrease.

1. **Summary of Findings:**

In this analysis, an exploration of the data highlights a mild negative correlation between income and convictions rate. This suggests that as income levels increase, there is a tendency for the convictions rate to decrease, though not very strongly.

1. **Insights:**

This analysis suggests that states with higher income levels are less likely to have higher conviction rates. This phenomenon may be attributed to the abundant resources available in these prosperous states, enabling individuals to access better legal representation and support, which in turn reduces the likelihood of convictions. Furthermore, it hints at a distinctive cultural behavior prevailing in affluent states, where a strong aversion to criminal activities is ingrained in the

****

**Project Title: Evaluating Optimal States for Health Tracker Smartwatch**

**Business Expansion**

societal fabric. Additionally, the presence of stringent government policies targeting corruption may loom large in these regions, instilling a palpable fear among individuals contemplating corrupt actions. As a result, the combination of resource abundance, cultural values, and strict anti-corruption measures converges to create a deterrent effect on criminal behavior, ultimately leading to lower conviction rates in high-income states."

1. **Conclusion in Relation to Business Expansion:**

Expanding into states with lower conviction rates, particularly high-income states, may offer a more favorable business environment, with potentially reduced legal and ethical risks. This can foster greater confidence in concluding business operations while maintaining a positive reputation.

However, it’s vital to acknowledge that selecting a state for business expansion is a complex decision. Other critical factors, such as tax rates, labor costs, and market demand, should also be carefully evaluated.

1. **Recommendations:**

* Evaluate Conviction Rates: Consider the conviction rates in potential expansion states, especially focusing on high-income states. States with lower conviction rates may offer a more favorable business environment, with potentially lower legal and ethical risks.
* Comprehensive Analysis: While conviction rates are important, conduct a comprehensive analysis that include factors like tax rates, labor costs, market demands, and regulatory environment. This holistic approach will help you make a well- rounded decision.
* Government Engagement: Engage with local government authorities to understand their policies and initiatives that may impact your business. Build positive relationship with regulatory bodies.

1. **Challenges Encountered During the Analysis:**

The challenges encountered during the analysis revolved around the limited dataset size and the exploration of the relationship between state income and state convictions. These constraints in data size posed difficulties in achieving statistical significance and comprehensive insights. Additionally, while investigating the connection between state income and state convictions, it became evident that the complex dynamics involved in this relationship required more extensive data and possibly supplementary variables to yield a deeper understanding.

1. **Recommendations for Future Research:**

* Expand Data Collection: To enhance the depth and breath of analysis, future research should aim to collect more comprehensive data. This might includes gathering additional variables related to law enforcement practices, socioeconomic factors, and cultural indicators within each states.

****

**Project Title: Evaluating Optimal States for Health Tracker Smartwatch**

**Business Expansion**

* Comparative Analysis: Extend the analysis to include a comparative study of different regions or countries to identify global trends and regional variations in the relationship between income and convictions.
* Longitudinal Analysis: Conducting longitudinal analysis over several years can provide insights into trends and changes in the relationship between state income and state convictions. This can help uncover evolving patterns and potential causal factors.