



# HEALTH TRACKER SMARTWATCH BUSINESS EXPANSION

Data Analytics Course Final Project

SUBMITTED BY:  
**ARSIE MAURICIO**

---

## PROJECT DESCRIPTION

---

A start-up company based in Los Angeles, California is launching a health tracker smartwatch into the market. The goal of the data analysis is to identify the state in the US with the best potential for business expansion using Data analytics tools such as SQL, Excel, Power BI, Python, and Business Metrics.

The key metrics used in this analysis are the following:

- *Health spending* — consists of average, minimum, and maximum spending for healthcare-related products and services per person.
- *Population size*- estimated number of people per state
- *Competitor companies' profit and expenditures*- consists of profit and spending data for companies
- *Property Prices*- average, minimum, and maximum property prices per square meter

Below is the project workflow that were performed in analyzing the data:

**1. Data Extraction and Transformation using SQL**

- Restore the PostgreSQL database `.tar` file containing economic indicators data.
- Use SQL to extract and transform data in the tables, create new table, and perform aggregations or calculations.

**2. Data Cleaning, Preparation and Analysis Using Excel**

- Import the data into Excel and perform additional data cleaning and preparation.
- Use datasets joined and extracted from PostgreSQL queries for the analysis
- Use pivot table, conditional formatting, and scatter plot for data visualization and analysis.

**3. Power BI Dashboard Creation**

- Perform DAX to create calculated columns and measures.
- Create a Power BI dashboard to visualize the key findings and insights from the data. This includes map chart, bar graphs, correlation matrix and other visualizations that highlight the most important economic indicators and trends.

**4. Data Analysis Using Python**

- Use the cleaned data to further analyze the data using Python libraries such as Pandas, Seaborn and Matplotlib to perform additional data verification, calculations, data visualizations, and exploratory data analysis.

**5. Calculate Business Metrics**

- Calculate Financial Viability Metrics such as Total Profit, Expenditures, Revenue, and Profit Margin to measure the health and stability of the business's finances.

**6. Report Writing**

- Write a report summarizing the findings, insights, and approach to analyze the data. The report also includes the overall recommendation, any limitations or challenges encountered during the analysis and recommendations for future research.

This section includes the tasks performed using each data analytics tool to achieve the data analysis goal.



→ **PostgreSQL**

*Task 1: Find the top 5 states with the total highest health spending based on the average and population.*

SQL CODE

```
SELECT hs.state_usa, avg_spending
FROM health_spending AS hs
LEFT JOIN population AS pop
ON hs.state_usa = pop.state_usa
GROUP BY hs.state_usa,
avg_spending
ORDER BY avg_spending DESC
LIMIT 5
```

Based on the SQL code above, below are the top 5 states identified with the total highest health spending. These states could have potential demand for health-related products.

	state_usa 	avg_spending 
1	Rhode Island	350.00
2	New Jersey	350.00
3	Massachusetts	350.00
4	Hawaii	350.00
5	Alaska	300.25

*Task 2: Find the top 5 states with the total lowest health spending based on the average and population.*

SQL CODE

```
SELECT hs.state_usa, avg_spending
FROM health_spending AS hs
LEFT JOIN population AS pop
ON hs.state_usa = pop.state_usa
GROUP BY hs.state_usa,
avg_spending
ORDER BY avg_spending ASC
LIMIT 5
```

Below are the top 5 states identified with the total lowest health spending. Healthcare spending is influenced by various factors. Some states may have lower spending but still provide high-quality care.

	state_usa character varying 🔒	avg_spending numeric 🔒
1	Idaho	125.50
2	Montana	125.50
3	Wyoming	125.50
4	Arizona	150.00
5	Mississippi	150.00

*Task3: Join the competitor dataset with the health spending dataset to see if there is any correlation between health spending per state and profit.*

#### SQL CODE

```
SELECT CORR(state_health_spending, total_profit) AS correlation
FROM(
SELECT hs.state_usa,
      ROUND(AVG(hs.avg_spending),2) AS
      state_health_spending,
      SUM(c.profit) AS total_profit
FROM health_spending AS hs
LEFT JOIN competitors AS c
ON hs.state_usa= c.state_usa
GROUP BY hs.state_usa
) AS health_profit
```

Based on the performed calculation using the SQL code, the correlation between health spending per state and profit is  $-0.25$ . This negative correlation means that higher health spending does not necessarily translate to higher profits. For example, the Rhode Island which was identified as one of the states with highest health spending has no generated profit while Wyoming with low health spending has generated a total profit of 16.7M.

→ **Excel**

The datasets joined and extracted from PostgreSQL queries are imported in excel for the data analysis. NULL values were replaced with 0 to clean the data.

*Task1: Create a pivot table to show the total health spending for each state based on the average and population.*

Below is the SQL code used to extract the data.

#### SQL CODE

```
SELECT hs.state_usa, avg_spending, estimate
FROM health_spending AS hs
LEFT JOIN population AS pop
ON hs.state_usa = pop.state_usa
GROUP BY hs.state_usa,
avg_spending,estimate
ORDER BY avg_spending DESC
```

Here's the sample pivot table to show the total health spending for each state.

Row Labels	Sum of avg_spending
Rhode Island	350
New Jersey	350
Hawaii	350
Massachusetts	350
Alaska	300.25
Connecticut	300
Pennsylvania	300
New Hampshire	300
Maryland	300
Virginia	300

*Task 2: Use conditional formatting to highlight the top 5 states with the highest health spending per capita.*

Row Labels	Sum of avg_spending
Rhode Island	350
New Jersey	350
Hawaii	350
Massachusetts	350
Alaska	300.25
Connecticut	300
Pennsylvania	300
New Hampshire	300
Maryland	300
Virginia	300

The top 5 states highlighted using conditional formatting in Excel has the same result with PostgreSQL.

*Task 3: Create a scatter plot to visualize the correlation between property prices and company profit.*

Below is the SQL code used to extract the data.

#### SQL CODE

```
SELECT pp.state_usa,  
       ROUND(AVG(pp.avg_price),2) AS  
       avg_prices,  
       ROUND(AVG(c.profit),2) AS avg_profit  
FROM property_prices AS pp  
LEFT JOIN competitors AS c  
ON pp.state_usa= c.state_usa  
GROUP BY pp.state_usa  
ORDER BY state_usa ASC
```

The extracted data was used to create a scatter plot to visualize the correlation of property prices vs. company profit.



The scatter plot shows a weak negative linear relationship. Using function CORREL() in Excel for this task, the correlation between property prices and profit is  $-0.14$ . It indicates that as property price increases, the profit is likely to decrease but the relationship is not very strong. Profitability depends on a combination of factors beyond just property prices.

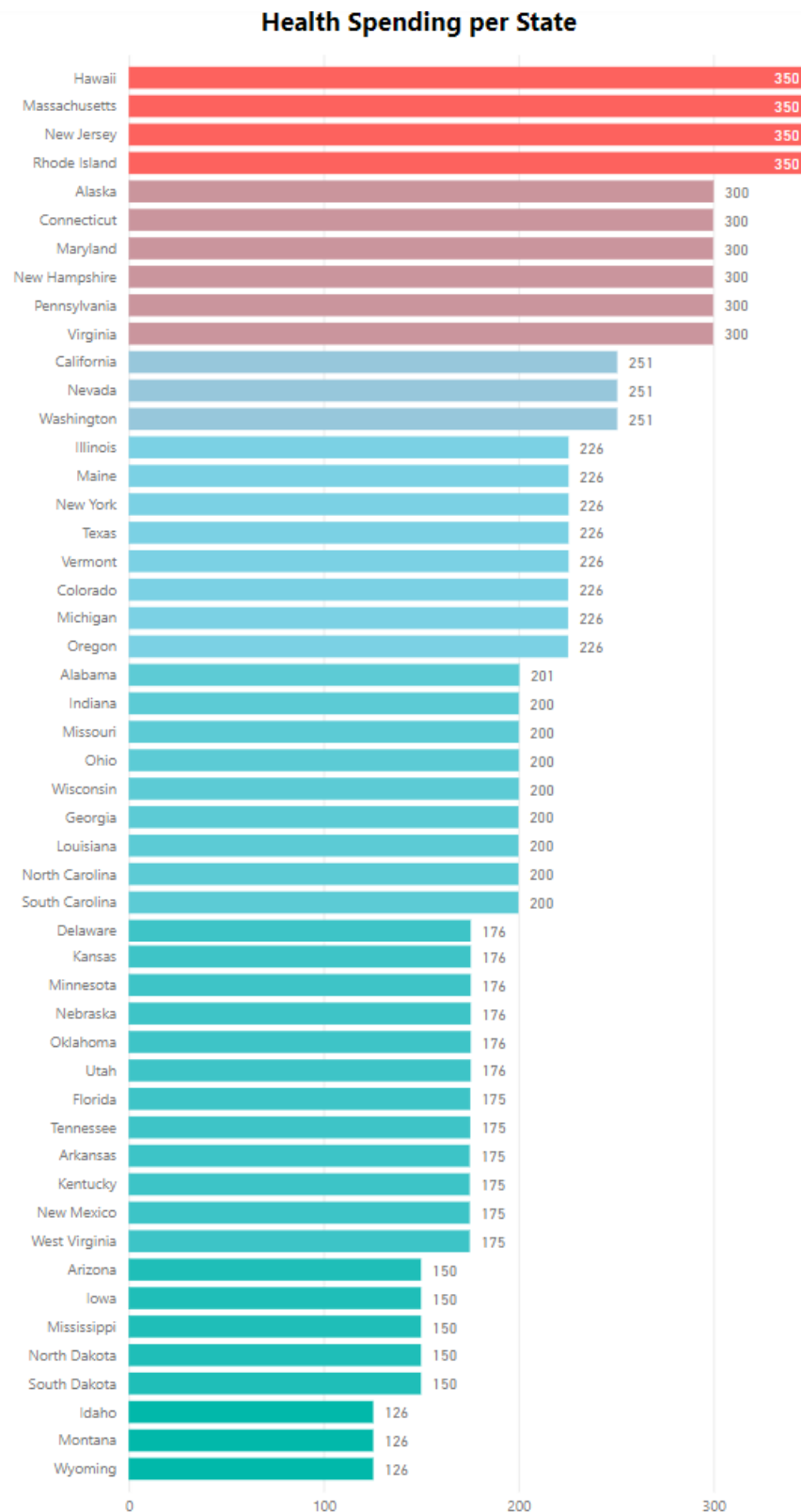
→ **Power BI**

Using PostgreSQL, we created new table to have a joined datasets of health spending, property prices, and profit for each state. This new table named *state combined data.csv* was imported to Power BI for data analysis and visualization. Below is the SQL code used to extract the data.

SQL CODE

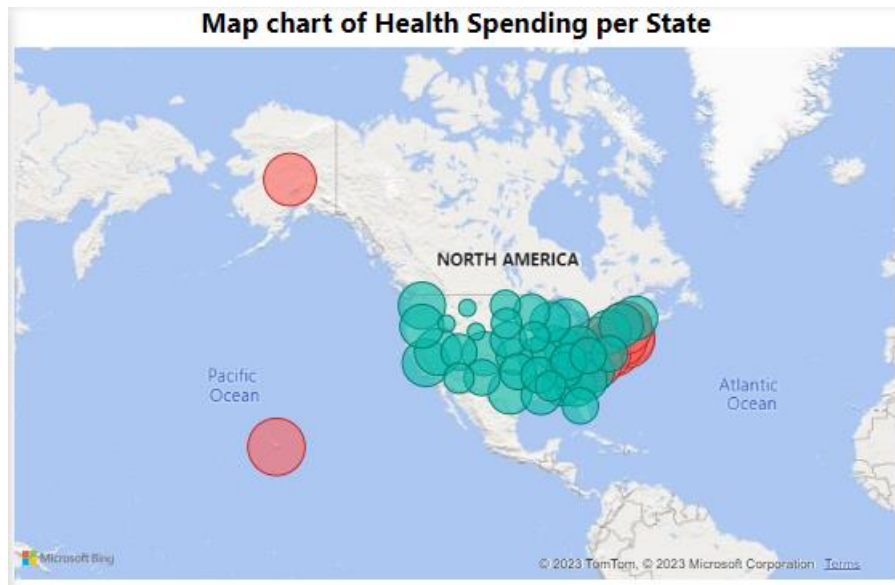
```
CREATE TABLE state_combined_data AS
SELECT
    hs.state_usa,
    ROUND(AVG(hs.avg_spending),2) AS
    avg_spendings,
    ROUND(AVG(pp.avg_price),2) AS avg_prices,
    ROUND(AVG(c.profit),2) AS avg_profit
FROM health_spending AS hs
LEFT JOIN property_prices AS pp
ON hs.state_usa= pp.state_usa
LEFT JOIN competitors AS c
ON hs.state_usa= c.state_usa
GROUP BY hs.state_usa
```

Task 1: Create a bar chart to show the health spending for each state.



The bar chart shows information about the average spending on health-related products in different states. Healthcare spending is influenced by various factors, including demographics, healthcare infrastructure, location, income levels, cost of living, lifestyle, insurance coverage, and state policies.

Task 2: Use the map visualization to highlight the top 10 states with the highest health spending per capita.



The bubble size highlighted in red shows the top 10 states with the highest health spending per capita.

Task 3: Create a correlation matrix to see if there is any correlation between property prices and profit.

In order to create a correlation matrix in Power BI, we created calculated columns using DAX to categorize the values of average price and average profit.

### DAX Formula

```
avg_profit_cat =
SWITCH(TRUE(),
    AND('state combined data'[avg_profit] >= 300000, 'state combined data'[avg_profit]
<= 600000), "300k to 600k",
    AND('state combined data'[avg_profit] >= 601000, 'state combined data'[avg_profit]
<= 900000), "601k to 900k",
    AND('state combined data'[avg_profit] >= 901000, 'state combined data'[avg_profit]
<= 1200000), "900k to 1.2M",
    'state combined data'[avg_profit] < 300000, "Below 300k",
    "Over 1.2M"
)
```

```
avg_price_cat =
SWITCH(TRUE(),
    AND('state combined data'[avg_prices] >= 1000, 'state combined data'[avg_prices] <=
2000), "1k to 2k",
    AND('state combined data'[avg_prices] >= 2001, 'state combined data'[avg_prices] <=
3000), "2001 to 3k",
    AND('state combined data'[avg_prices] >= 3001, 'state combined data'[avg_prices] <=
4000), "3001 to 4k",
    'state combined data'[avg_prices] < 1000, "Below 1k",
    "Over 4k"
)
```



avg_profit_cat	1k to 2k	2001 to 3k	3001 to 4k	Over 4k	Total
300k to 600k		1		1	2
900k to 1.2M	1	2			3
601k to 900k	3	1			4
Below 300k	5	1	2	2	10
Over 1.2M	14	12	3	2	31
<b>Total</b>	<b>23</b>	<b>17</b>	<b>5</b>	<b>5</b>	<b>50</b>

The correlation matrix shows the count of states within specific profit categories and property price categories. Highlighted in red shows that majority of the states with property prices between 1k to 3k have a higher count of states where businesses are potentially more profitable which generated over 1.2M. States with both high and low average profits can be found across a range of property price categories. This implies that profitability does not solely depend on property prices, and other factors may influence a state's profit levels. Lower property prices may be more favorable for business profitability.

→ **Python**

#### Importing Python Libraries

For us to analyze the given data, we first have to import the necessary Python libraries.

#### PYTHON CODE

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

#### Dataset Loading

For this task, we're going to first mount the drive then load up the .csv file named *state combined data.csv* into a pandas dataframe and store it in a variable named *data*.

#### PYTHON CODE

```
from google.colab import drive
drive.mount('/content/drive')

data = pd.read_csv('/content/drive/MyDrive/Refocus Final
Project/state combined data.csv')
```

To check if the loading was successful, we can use the `data.head()` to show the top 5 rows in the data.

	state_usa	avg_spending	avg_prices	avg_profit
0	Alabama	200.50	1797.50	1138459.39
1	Alaska	300.25	2684.00	1969126.59
2	Arizona	150.00	2356.75	2942282.99
3	Arkansas	175.00	1499.25	743958.99
4	California	250.75	5832.50	398315.70

*Task 1: Use Pandas to find the top 5 states with the highest health spending per capita.*

#### PYTHON CODE

```
data_sorted = data.sort_values(by='avg_spending', ascending=False)
top_5_states = data_sorted.head(5)[['state_usa', 'avg_spending']]
top_5_states
```

	state_usa	avg_spending
10	Hawaii	350.00
38	Rhode Island	350.00
29	New Jersey	350.00
20	Massachusetts	350.00
1	Alaska	300.25

Using Pandas, the result shows the top 5 states with the highest health spending per capita. Same result with the data analytics tools illustrated above.

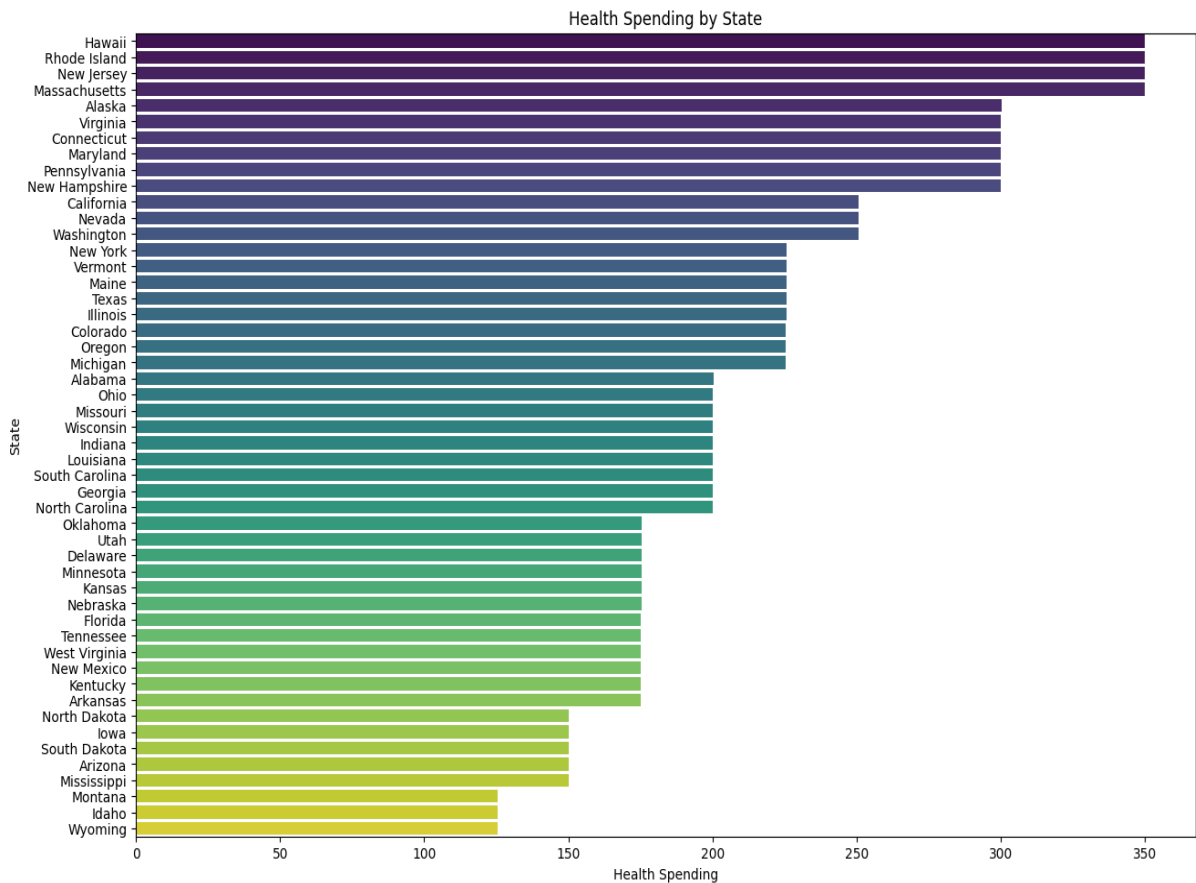
*Task 2: Use Matplotlib and/or Seaborn to create a bar chart to show the health spending for each state.*

#### PYTHON CODE

```
plt.figure(figsize=(15, 10))
sns.barplot(x='avg_spending', y='state_usa', data=data_sorted,
           palette='viridis')

plt.xlabel('Health Spending')
plt.ylabel('State')
plt.title('Health Spending by State')

plt.show()
```



The bar chart shows the health spending for each state sorted by highest to lowest.

*Task 3: Use Matplotlib and/or Seaborn to create a scatter plot to visualize the correlation between property prices and profit.*

### PYTHON CODE

```
plt.figure(figsize=(8, 5))

sns.scatterplot(x='avg_prices', y='avg_profit', data=data_sorted,
color='green', alpha=0.6)

plt.xlabel('Average Property Price')
plt.ylabel('Average Profit')
plt.title('Correlation of Property Prices vs. Profit')

plt.show()
```



The scatter plot we created using Matplotlib/Seaborn shows the same result that we got using Excel. It shows that there is a negative correlation between property prices and company profit. We can observe that there's a concentration of data points in less than 3k property prices and above 1.2M average profit.

#### PYTHON CODE

```
correlation_coefficient =
data['avg_prices'].corr(data['avg_profit'])

print("Correlation Coefficient between Property Prices and Company
Profit:", correlation_coefficient)
```

The Python code was used to verify the correlation result in this task which also resulted to  $-0.14$ .

#### → **Business Metrics**

The competitor companies' dataset consists of profit and expenditures data. Based on the available data, we were able to calculate the Financial Viability Metrics such as Total Profit, Expenditures, Revenue, and Profit Margin. This information can be used to assess profitability in different states.

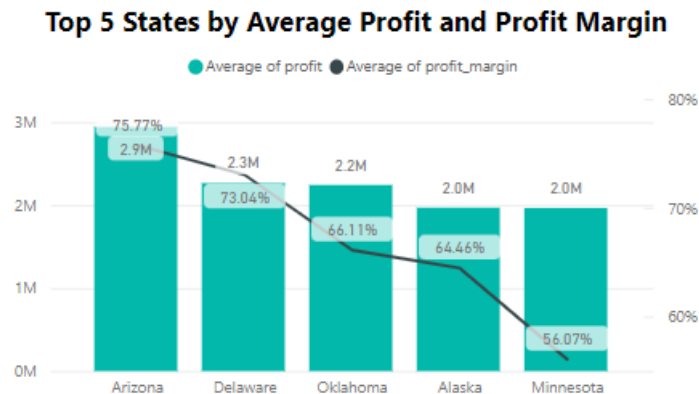
#### Financial Viability Metrics

Total Expenses= SUM of research & development, administration, and marketing spent

Revenue= SUM of Profit and Total Expenses

Profit Margin= (Profit/Revenue) \*100

Using Power BI, we were able to create new measures to calculate the Total Profit, Total Revenue, and Net Profit Margin as well as to visualize the top 5 highest states based on the Average Profit and Profit Margin.



Based on the combination chart, these states have relatively high average profits and high profit margins. It indicates that businesses in these states are generating substantial profits and are efficient in converting revenue into profit.

Among other states in US, Arizona has the highest average profit of 2.9M and profit margin of 75.77%. These metrics for Arizona suggest that businesses in the state, on average, generate a substantial amount of profit, and their profit margins are also relatively high.

---

## CONCLUSION

---

→ **Summary of Findings**

- **Average Health Spending**

*Highest:* The top five states with the highest average spending on health-related products are Rhode Island, New Jersey, Massachusetts, Hawaii, and Alaska. These states have an average spending of 300.25 to 350.

*Lowest:* The top five states with the lowest average health spending are Idaho, Montana, Wyoming, Arizona, and Mississippi ranging from 125.50 to 150.

- **Correlations**

*Health Spending vs. Profit:* There is a weak negative correlation (-0.25) between health spending per state and profit.

*Property Prices vs. Profit:* There is also a weak negative correlation (-0.14) between property prices and company profit. The correlation matrix indicates that there are variations in profit and property prices across different average profit and average property price categories. States with average profits categorized as "Over 1.2M" are predominantly in the "1k to 2k" followed by "2001 to 3k" average price category.

- **Profitability**  
States like Arizona, Delaware, and Oklahoma have relatively high average profits and profit margins. Alaska and Minnesota also show promising profitability metrics.

→ **Insights**

- **Average Health Spending**

*Highest:* States with higher average spending may indicate a stronger demand for health-related products, including health tracker smartwatches.

*Lowest:* While these states have lower health spending, other factors should still be considered and may still make them attractive markets.

- **Correlations**

*Health Spending vs. Profit:* The weak negative correlation (-0.25) between health spending per state and profit means that the higher health spending does not necessarily lead to higher profits.

*Property Prices vs. Profit:* The weak negative correlation (-0.14) between property prices and company profit indicates that lower property prices may be associated with higher profitability. However, other factors and context should be considered since it shows a weak relationship.

- **Profitability**

The States identified with higher profit and profit margin have already demonstrated potential for profitability based on competitors' data. They could be potential candidates for business expansion.

→ **Answer to Business Question**

If the goal of the analysis is to determine which state to select for business expansion, our findings suggest that prioritize states with high average profits and profit margins, such as Arizona for expansion efforts. Despite being one of the states with lowest health spending, Arizona still shows a favorable average profit and profit margin indicating healthy financial performance. Focus on markets where your competitors have been successful. While there is a negative correlation between health spending and profit, states with both high health spending and profitability should still be considered for expansion such as Alaska which is both included in one of the top 5 states with highest health spending and highest average profit. However, consider conducting further market research to understand specific consumer preferences and market dynamics in the selected states before making expansion decisions.

→ **Recommendations**

- **Target Profitable Markets:** If your goal is to expand your business into states with higher profit potential, consider focusing on states in the "Over 1.2M" profit category. These states show a mix of property price categories, but they offer the highest profit potential. These states have shown profitability for your competitors. When identifying a state for business expansion, it's generally recommended to focus on average profit. Average Profit considers Per-Unit Performance typically per state in this case. This means it takes into account the efficiency and effectiveness of your operations in generating profit relative to the size or population of the state.
- **Monitor Profit Margins:** Pay attention to states with favorable profit margins, as they may indicate higher pricing flexibility or cost efficiency.

- **Evaluate High-Health Spending States:** Analyze the potential for market demand in these states.
- **Assess Property Prices:** Consider states with a balance of profitability and favorable property price trends. A favorable property price can lead to cost savings. States with property prices between 1k to 3k have a notable number of businesses and could potentially be profitable markets.
- **Consider Population:** Take into account the population of each state. States with larger populations offer a larger customer base but may also have more competition. Smaller states with high average spending and profit margins may provide a better entry point. Entering markets with less competition may provide opportunities for growth.
- **Identify Competitive Advantage:** Identify your competitive advantages, such as product features, pricing, or marketing strategies, to differentiate your smartwatches in the chosen markets.
- **Consider Pilot Programs:** Implement pilot programs in selected states to test the market response and better understand the dynamics between property prices, profit, and consumer behavior in those areas.

→ **Challenges Encountered during the Analysis**

- **Limited Data Availability:** Data such as date, regions, investment cost, and different healthcare spending categories are not provided. These data can be used to observe trend over time period, regional trends, and calculation of ROI, while categorization simplifies the analysis and makes it easier to communicate findings and recommendations to stakeholders. Also, type of profit is not well defined in the description if it is net or gross profit.
- **Data Integration:** Combining data from various datasets and formats into a single dataset can be time-consuming and error-prone. Ensuring data consistency and compatibility between different datasets is crucial.
- **Choosing Relevant Variables:** Identifying the most relevant variables (e.g., health spending, profit, property prices) and determining their impact on business expansion can be complex.

→ **Recommendations for Future Research**

- **Market Research:** Conduct thorough market research in the selected states to understand consumer preferences, demographics, and competition. Identify the specific needs and preferences of the target audience in each state.
- **Competitor Analysis:** Conduct a comprehensive analysis of competitors in the health tracker smartwatch industry. This should include a review of their product offerings, market share, pricing strategies, and customer reviews. Understanding the competitive landscape can help identify opportunities for differentiation.
- **Long-Term Profitability:** Evaluate the long-term profitability of expanding in the identified states. Analyze the states with the highest net profits to understand what factors contribute to their success. Consider factors such as potential growth rates, ROI, sustainability of healthcare spending trends, and the scalability of your business model. This analysis can assist in making informed decisions about resource allocation.
- **Customer Segmentation:** Segment your target audience within each state based on demographics, health spending habits, and other applicable metrics. Develop targeted marketing campaigns to address the unique needs of each segment.

- ***Data Quality and Availability:*** Find relevant and up-to-date data for all necessary variables, especially for competitor data.
- ***Income and Corruption Data Analysis:*** Consider to analyze other key metrics such as state income and corruption levels for each state.