Austin Rippee
April 25th, 2023
CSIS 586

Final Project

Project Video: https://youtu.be/wL8kxxMyV8U

Problem Description:

The dataset that I chose is data pertained to a specific runner over the course of a year. I want to find out how the heart rate and cadence changed as the altitude increased. Then, I want to find out if the speed of the runner changed as the altitude increased and if that influenced the distance that was run.

Description of Data:

Source of the data and data format: https://www.kaggle.com/datasets/mcandocia/running-heart-rate-recovery?select=similarities.csv (In CSV format)

How the data was obtained: Through a Kaggle download including six CSV files.

Was the data used as it is? Or has the cleaning been done? The data was used as is. I had to upload it manually through SQL queries but did not have to clean up the data as many columns already had the data, I needed to use to make my conclusion.

Size of the data: According to Kaggle, the download is 18MB, but the data explorer says 63.96MB.

Analysis Techniques:

Tools used: I used the built-in tools associated with GCP: BigQuery, SQL workspace, gcloud, and gsutil. Using these tools to create the bucket and instance and connected them. Once everything was setup, I was able to use the BigQuery SQL workspace to be able to find the solutions that I needed then I used Google Sheets to create visual graphs for the data to draw conclusions.

Computing resources: has the analysis been done on your local machine or on cloud? I used the cloud using Google Cloud Platform.

How did you analyze the data? I analyzed the data by using SQl queries to simplify and manipulate the data to draw a solidified conclusion.

Conclusion:

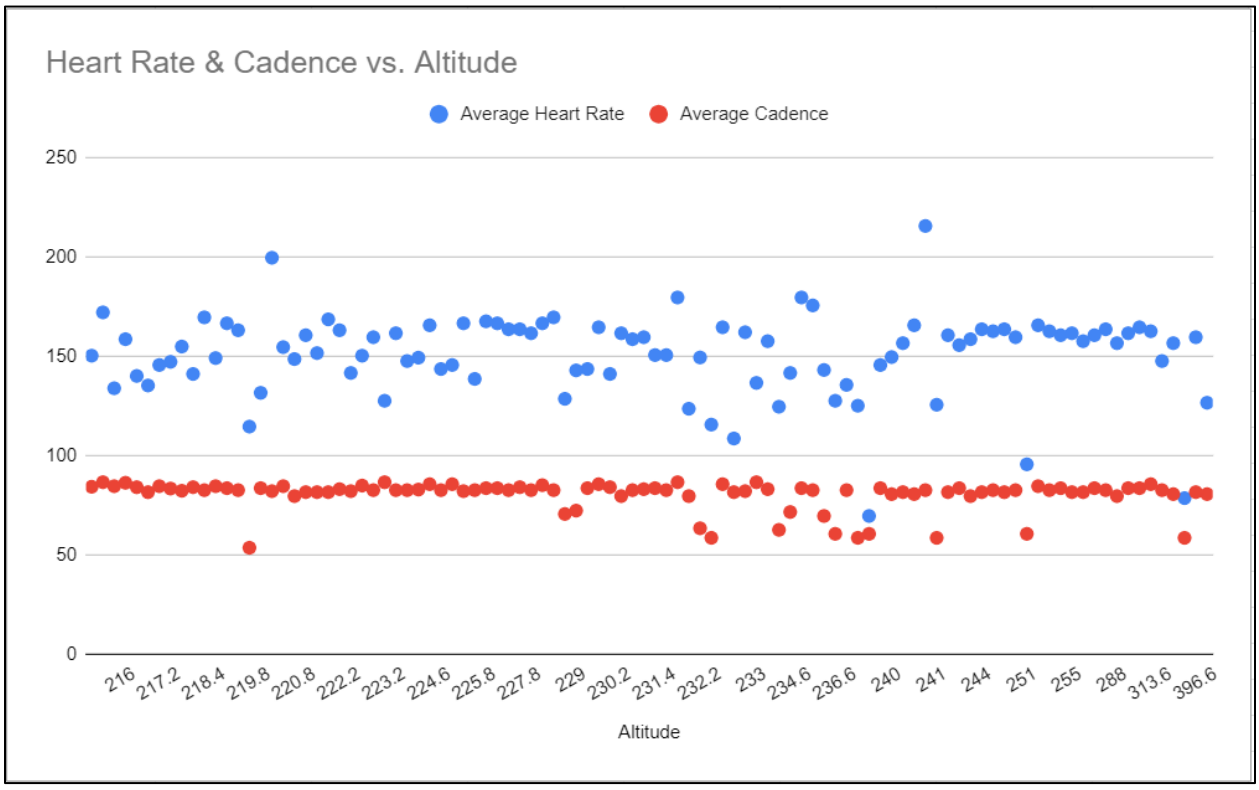
To figure out both of my problems as stated in the problem description, I needed to use two SQL queries to do so. I was able to use a join statement to combine the two separate problems I wanted to find a solution for. The following is the SQL query I used in my problem:

```
SELECT ROUND(AVG(a.heart_rate),1) as `Average Heart Rate`,
ROUND(a.altitude,1) as `Altitude`,
ROUND(AVG(a.cadence),1) as `Average Cadence`,
ROUND(AVG(b.distance),1) as `Average Distance`,
ROUND(AVG(b.speed),1) as `Average Speed`
FROM `csis586-final.finaldataset.s1_laps_summary` a

JOIN (SELECT altitude, AVG(distance) as distance, AVG(speed) as speed FROM `csis586-final.finaldataset.s1_laps_summary` GROUP BY altitude) b
ON a.altitude = b.altitude

GROUP BY a.altitude
HAVING COUNT(a.heart_rate) = COUNT(a.cadence) AND COUNT(b.distance) = COUNT(b.speed)

ORDER BY a.altitude DESC
LIMIT 100;
```

PROBLEM #1

The first problem was that I wanted to find out how the heart rate changed as the altitude increased. More specifically, when the altitude increases, there is less oxygen in the atmosphere leading to more energy needed to keep the same pace. The second problem I wanted to find out was how the cadence changed as the altitude increased. More specifically, I wanted to see if a higher altitude can cause the runner to slow down. After running through the query, I was able to use a scatter-chart to have the best visualization of this data.

The chart with the points of data is shown below:

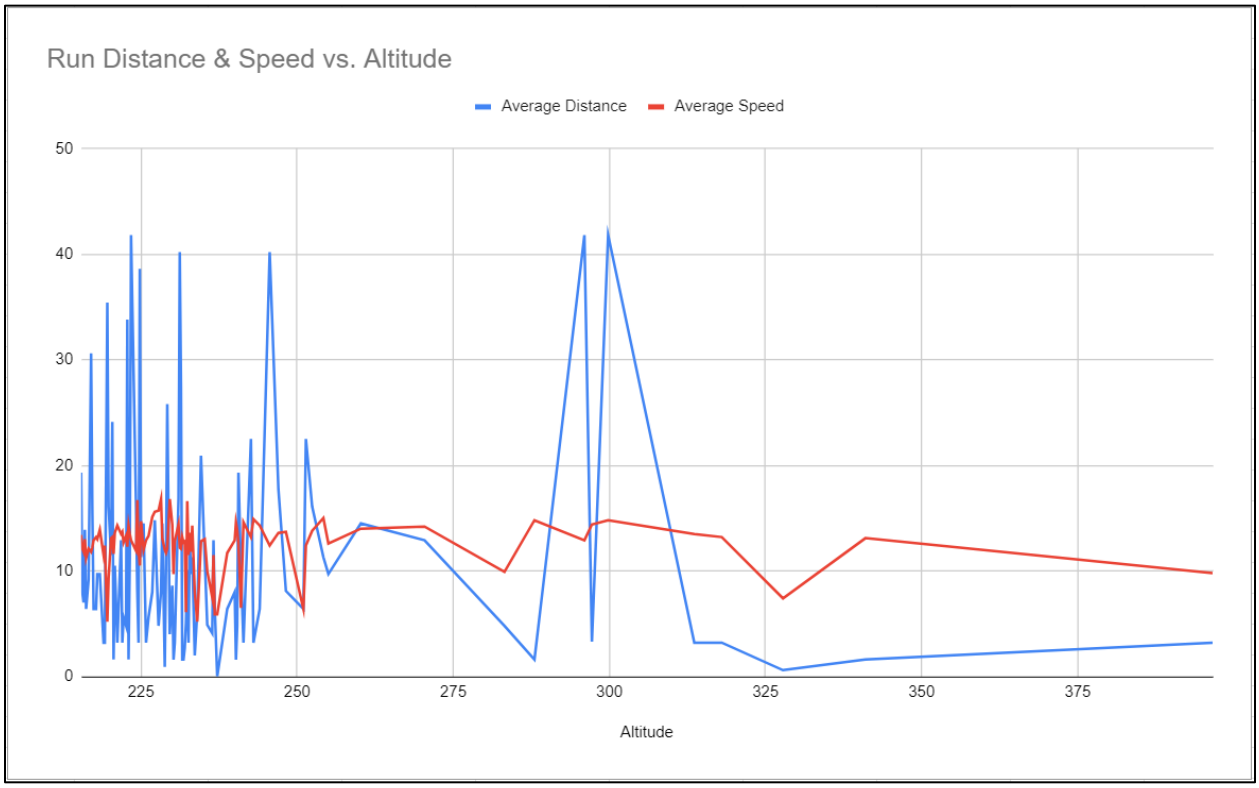


After looking through the graph, I determined, for the first problem, that the average heart rate as the altitude increases, becomes a lot more sporadic. When the altitude is lower, the heart rate tends to stay even sitting around 125-170 BPM with a 145 BPM average. There were two outliers of 110 and 200 BPM. But when the altitude is higher however, the data points become a little more scattered with an average of about 160 BPM. The outliers as the altitude increased, were around 70, 80, 95, and 220. Therefore, we can conclude that altitude does not necessarily increase the heart rate but makes it not as consistent per run. Lastly, I determined based on the data of the graph, the cadence does not change as the altitude increases. This proves that even though the energy produced by the runner goes up, their form does not deteriorate.

PROBLEM #2

The second problem I wanted to find out was if the speed of the runner changed as the altitude increased and if that influenced the distance that was run. More specifically, I wanted to see if an increase in altitude causes the runner to slow down and if the runner chose to run a shorter distance since it is harder to breathe at a higher altitude. After running through the query, I was able to use a bar graph to have the best visualization of this data.

The chart with the points of data is shown below:



After looking through the graph, I determined, for the first problem, that as the altitude increases, the speed tends to not have an effect. This shows that the speed of the runner is roughly around the same as when the runner was at a lower altitude. Though, it is worth noting that there were more frequent runs at a lower altitude level than there were at a higher altitude. Lastly, I determined from the graph that the average distance changed dramatically as the altitude increased. You can see on the graph that after 300 feet of altitude, the average distances of the run were very short compared to that below 300 feet. Therefore, we can conclude that altitude has a large effect on the distance run.