**DAT 210 Final Project: Written Report for RfG**

**By Alexander Varljen**

**June 15, 2022**

**Table of Contents**

1. **Introduction: Page 3**
   1. **Situation and Problem Analysis: Page 3**
   2. **Research Questions: Page 4**
   3. **Additional Remarks: Page 5**
2. **Data Evaluation: Page 5**
   1. **Sources (1-11): Page 5**
   2. **Links to Data Sources: Page 10**
3. **Data Processing Steps: Page 12**
   1. **Step Outline: Page 12**
   2. **Step Importance: Page 13**
   3. **Usage Regulations: Page 14**
4. **Data Analysis: Page 15**
   1. **Analysis Concepts: Page 15**
   2. **Recommended Solutions: Page 15**
5. **Concluding Remarks: Page 16**
6. **References: Page 17**

## Introduction

• Situation and Problem Analysis

Before further research is conducted, we should analyze the current situation for Running for Glory. I will separate this analysis into three categories: the current facts about the organization, the plans and expectations of the organization, and current limitations. First, in terms of the current facts about RfG’s situation, RfG is a fitness retailer that has a dedicated customer base and one outlet. It provides a specific service, which is centered around running, and maintains inventory records and can readily produce basic sales data. In terms of RfG’s plans and expectations, RfG’s management would like to expand into the suburbs to serve an additional customer base, which is younger. As far as limitations to further research and analysis are concerned, the only actionable data that RfG has on site is sales and inventory records. However, RfG seeks to have more data pertaining to suburban demographics so that it can develop a strategy it can use to expand into the Seattle suburbs.

We will define the problem in two steps: first generally and then with more specific detail. The general problem is that, while RfG would like to expand its customer base to include a younger demographic, their sales don’t go beyond the current customer base. The details pertaining to this problem include having only one outlet, the lack of expansion into the suburbs where new and younger potential customers can be found, and a lack of current actionable data related to the target demographic.

It seems that there will be three categories of business processes that will be affected by the problem identified. One category is that of Management process. The RfG Management process that are likely to be affected are finding locations for new stores as RfG expands into the suburbs and developing a business strategy. Another category is that of Operational processes. The RfG Operational Processes that could be affected include marketing and sales, as RfG will be seeking out new customers, and budgeting, as business expenses will change. Another category is that of supporting processes. The RfG supporting processes that will likely change include recruiting, as expanding the business will require new employees, and customer support, as potential new customers may not be familiar with the business and products.

• **Research Questions**

* The following are research questions related to RfG’s business problem:
* What Seattle suburban locations would be most viable for a new RfG outlet location?
* What is the expected market size for potential customers who are 35 years of age and younger?
* How is the athletic equipment market going to change in terms of sales over the next decade?
* **Additional Remarks:**
* Running for Glory has maintained itself and a dedicated customer base for over forty years. However, it now wishes to expand to serve a younger demographic. Some challenges it faces include gathering more data, adding new outlets, and making new adjustments to its Management, Operational, and Supporting processes. Until further analysis is conducted, it may be helpful to delve into the literature of research conducted by those who have already analyzed similar situations. One example is a study conducted by Comber (2021), which discusses the opportunities that Big Data provides for retail. Another example is a study conducted by Zhang (2021), which discusses the use of Big Data sources to find out how much private households in each area are likely to spend in retail. Sources like these, as well as further research and analysis of Running for Glory’s specific situation, will help our consulting company aid RfG in coming up with a viable solution to their business problem and future opportunity.

### Data Evaluation

In this section, each data source will be discussed in terms of its type, its applicability to RfG’s business problem, its quality, and its limitations (note that, after the title of each source, there is a number that corresponds to its digital link in the “Links to Data Sources” section).

* **Source 1:**

The first data source is called “Washington Business Tax & Premium Database – Tax and Premium Statistics” (1) and outlines the tax rate that a business located in Washington will have. This source is relevant to RfG’s needs, as RfG will have to accurately calculate business expenses as it expands. One example of the applicability of this data source is that one of the charts breaks down the total annual cost as a percentage of annual revenue of a Washington business by size in Washington. Some notes about the quality of the data include that it comes from official Washington State records, that the methodology is given in a separate report, and that the data is specific to the tax and insurance premium rates of Washington State businesses. Some limitations include that the updates to the data are slow, with this current report being from 2009, and that, while analyzed data is readily available, individual data points are not.

* **Source 2:**

The next data source is called “Washington Income Distribution Update (2013)” (2), which outlines the income of Washington residence. The relevancy of this source is that it can help RfG determine the availability of spending money those new potential customers may have, when paired with other information, such as Washington or Seattle living expenses. One example of the applicability of this data source is a chart which displays the incomes of Washington residents, separated into deciles, which are compared to the federal poverty level. Some notes on the quality of this source include that it is derived from official Washington State records, and that the data source is a breakdown of aspects of Washington State income. One limitation is that the data was last gathered in 2011. Also, while analyzed data is readily available, individual data points are not.

* **Source 3:**

The next data source is called “Seattle Office of Economic Resources – Occupational Dashboard” (3), which is an interactive chart which displays occupational data on Seattle workers. The relevancy of this source is that it will help RfG plan for wage expenditures as it hires new employees for its future suburban outlet. An example of the applicability of this data source is that, by using the option bar above the graph, one can get the average hourly earnings of a Seattle resident by occupation. Some notes on the quality of this source include that it is gathered by an official Seattle city government office, that it is updated annually, and that the raw data can be downloaded, checked, and analyzed. There seem to be few apparent limitations to the retrieval of this data. One limitation to this data source is that there doesn’t seem to be an explanation of the methodology of how this data was gathered.

* **Source 4:**

The next data source is called “FRED Economic Data – Consumer Price Index for All Urban Customers – Monthly, Not Seasonal Adjusted: Seattle-Tacoma-Bremerton, WA (Release Table)” (4), which gives various cost of living expenses in the Seattle-Tacoma-Bremerton area. The relevancy of this source is that it can help determine the price that customers typically pay for goods and services, as well as what some business costs RfG will likely face as they add a new outlet to their company. An example of the applicability of this source is the CPI for electricity, which is one aspect of an operating cost for RfG. Some notes on the Quality of this data are that it is gathered by the economic research organization FRED, that it is regularly updated, and that it can readily be placed in a chart using the websites tools. There seem to be no apparent limitations to the retrieval of this data. One of the limitations of this data is that an analyst may need knowledge of economic terms, like CPI, to understand this data source.

* **Source 5:**

The next data source is “SAIDI Electric Reliability Data (2014-2017)” (5), which gives the number of minutes that a power outage has lasted for each month in Seattle from 2014, onward. The relevancy of this data is that it gives a secondary consideration for RfG’s business plans going forward. An example of the applicability of this source is that RfG can see for how many minutes power is likely to go out for a given month in their new suburban outlet, based on past records, and then plan accordingly. Some notes on the quality of this data include that it is gathered and displayed by the Seattle Open Data Program, that it is updated monthly, that it can be downloaded and analyzed, and that one can use the websites visualization software with this data set. There seem to be no apparent limitations to the retrieval of this data. One of the limitations is that the average is for the city area and can’t, for instance, help one differentiate between the average time of power outages in urban versus suburban Seattle.

* **Source 6:**

The next data source is “SPD Crime Data: 2008 to Present” (6), which gives the date, location, and the specific offense of all crime reports in Seattle from 2008 onward. The relevancy of this source is that it can help RfG plan for security needs in its new outlet location. An example of the applicability of this source is that, once a location is selected for the new suburban outlet, one can analyze this source for the crime rate of that location and, specifically, any crimes that are likely to affect a business, such as robbery. Some notes on the quality of this source include that the data was gathered by the Seattle Open Data Program, that it is updated every few days, and that there are multiple columns for location (including the exact latitude and longitude of the incident), offense, and date of report. There seem to be no apparent limitations to the retrieval of this data. One of the limitations of this data source is that response time doesn’t seem to be included in the data set.

* **Source 7:**

The next data source is called “Building Permits” (7) and displays the permit status of Seattle buildings. The relevancy of this source is that, if RfG is planning to purchase a completed building for their new outlet in the suburbs, they can check the permitted status of the building from this data set. One example of the applicability of this source is that a building being considered for purchase can be looked up on the source by its address to see whether any or all the necessary building permits are obtained for said building. Some notes on the quality of this source include that the data was gathered by the Seattle Open Data Program, that it is updated daily, and that there are multiple columns for data points ranging from location of the building to the type of permit that was applied for. There seem to be no apparent limitations to the retrieval of the data presented. One of the limitations of this source is that several columns, such as the column for permit expiration dates, are empty.

* **Source 8:**

The next data source is called “Statista – Share of joggers who went long distance jogging or running in 2021, by generation” (8), which gives the percentage of long-distance runners and joggers in the U.S. by generation. The relevancy of this source is that it can allow RfG to get a general idea of how big the market share of younger adult runners there are, as they are planning to expand to the suburbs. One example of the applicability of this source is comparing the proportion of millennial and gen z runners to gen x runners. Some notes on the quality of this source include that it was collected from an online survey with over one-thousand participants and that this survey was given in 2021. Some limitations of this source include that an in-depth analysis of the data requires a subscription and that the range of subjects include participants from across the United States, rather than just the State of Washington or the City of Seattle.

* **Source 9:**

The next source is called “Fitness Equipment Market Type (Cardiovascular Equipment, Strength Training Equipment, and Others), End User (Home Customer, Health Club/ Gym, and Other Commercial User): Opportunity Analysis and Industry Forecast” (9), which is an analysis of the growth of the fitness equipment market over the next eight years. The relevancy of this source is that it can aid RfG in forecasting how much the market for their new outlet can grow over time. One example of the applicability of this source would be comparing the running equipment sales of 2020 to their projected sales in 2028. Some notes of the quality of this source include that its data and analysis were conducted by a professional market analysis company. Some limitations of this source include that this data and report needs to be purchased and that, while running equipment is included in the source, most of the source includes data and analysis of other areas of the fitness market, which may not be applicable to RfG.

* **Source 10:**

The next source is called “Multi-Use Trails (Seattle Only) (data table)” (10), which gives data on the pedestrian and bike trails in Seattle. The relevancy of this source is that it provides information on common running areas in Seattle, which RfG can use to provide information to new customers as part of their customer service or, possibly, use to pick a location for their new suburban outlet. An example of the applicability of this source is that the locations of free and open, nearby pedestrian trails can be given to a customer who is new to running, or the area. Some notes on the quality of this source include that it is updated every few days, that the data table can be downloaded for analysis, and that the source comes with a map of the trails. There seem to be no apparent limitations to the retrieval of the data available. One of the limitations of this source is that there are no pictures of the trails themselves, which would likely be of interest to runners.

* **Source 11:**

The next source is called “Sensys Traffic Volumes (Vol Avg \* APEG)” (11), which gives daily Seattle traffic data by location. The relevancy of this source is that RfG employees will have to drive to and from the new outlet’s location, and that potential customers will drive past the store location in specific volumes, depending on the street location. A potential application could be to see what the average traffic volume of a potential outlet location is, and, therefore, get a general idea of how many drivers will see the new location while commuting. Some notes on the quality of this source include that it is from the Seattle Open Data Program, that it is updated daily, that the source of the data are Seattle traffic sensors, and that the sensor location, number of vehicles that passed, and date of recording are all part of the data set. In terms of limitations to this source, there doesn’t seem to be any restriction on the downloadability and analysis of this source. One thing to keep in mind, however, is that this source changes from day to day, albeit as additions to the already existing data set.

* **Links to Data Sources:**

(1) <https://ofm.wa.gov/sites/default/files/public/legacy/economy/business_tax/tax_and_premium_statistics.pdf>

(2)

<https://ofm.wa.gov/sites/default/files/public/legacy/economy/income_wealth/income_wealth_update_2013.pdf>

(3)

<https://www.seattle.gov/office-of-economic-development/workforce-development/data-and-resources>

(4)

<https://fred.stlouisfed.org/release/tables?rid=10&eid=36229#snid=36238>

(5)

<https://data.seattle.gov/City-Business/SAIDI-Electric-Reliability-Data-2014-2017/ahy7-fn62>

(6)

<https://data.seattle.gov/Public-Safety/SPD-Crime-Data-2008-Present/tazs-3rd5>

(7)

<https://data.seattle.gov/Permitting/Building-Permits/76t5-zqzr>

(8)

<https://www.statista.com/statistics/227423/number-of-joggers-and-runners-usa/>

(9)

<https://www.alliedmarketresearch.com/fitness-equipment-market#:~:text=The%20global%20fitness%20equipment%20market%20size%20was%20%2411.6,and%20monitoring%20devices%20required%20for%20various%20physical%20exercises>

(10)

<https://data-seattlecitygis.opendata.arcgis.com/datasets/SeattleCityGIS::multi-use-trails-seattle-only/explore?location=47.606920%2C-122.326650%2C11.32&showTable=true>

(11)

<https://data.seattle.gov/Transportation/Sensys-Traffic-Volumes-Vol-average-APEG-/9y6s-jz9z>

### Data Processing Steps

• Step Outline

Figure 1: Data Process Steps Flowchart:



The **Figure 1** flowchart outlines my data processing steps. The first step of my data processing strategy is to separate the secondary and primary sources of data, as the secondary sources of data that I collected can only be checked for quality and relevance. The second step is the validation process, where all the data sources are checked for their relevance to the current business problem, as well as their quality. The second step is to check the primary data sources for a common variable, like location, and if a common variable is found, harmonize the data sources. An example of where this could be important, would be in the case of some of my sources, like *SPD Crime Data: 2008 to Present* (2022) and *SAIDI Electric Reliability Data (2014-2017)* (2022), where there are variables for location that could be used to harmonize the data sets, which would mean that there would be one data set to find the crime rate of a potential new outlet location for RfG, the rate of power outages for that location, and other variables from other like sources of data.

The third step would be editing the primary data sources. This would include correcting inconsistent data, omissions of unhelpful data, and dealing with outliers appropriately (Bougie, R. & Sekaran, U., 2020). An example of where this could be used is in the case of one of my data sources, which is *SPD Crime Data: 2008 to Present*, which has a few columns that are unfilled (Seattle Open Data, 2022). The fifth step is coding, which is the process of entering data in a way that the data is simpler and numerical, which is usually handled during data collection (Bougie, R. & Sekaran, U., 2020). An example of this could be giving an integer for each type of offense in the *Crime Data: 2008 to Present* data set (Seattle Open Data, 2022). The sixth and final step is tabulation, which would entail arranging or rearranging data sets such that they are more available for access. Much of my current sources of data have no need of further tabulation. However, if new primary data is gathered, or a survey, interview, or another data gathering technique is to be employed on RfG’s behalf, the tabulation step may need to be used.

• Step Importance

* I will go over my data processing systems relevance to RfG’s business problem, step by step. The first step, which is separating the primary from the secondary sources is important, as I can’t edit, manipulate, or otherwise change the data points of my secondary data sources. The second step, which is validation, is necessary to check if RfG could benefit from a given source. The third step of harmonizing, in practical terms, would allow me to make simpler data sets that combine the variables that could aid RfG in business operations decisions, such as choosing a good outlet location. The fourth, fifth, and sixth steps, which are editing, coding, and tabulation, would make the data easier to use and allow one to get useful results while performing analysis.

• Usage Regulations

There are a few ways to ensure the integrity of the data sources that will be used to help solve RfG’s business problem. One way is to, once the data is collected, validated, harmonized, coded, edited, and tabulated, guard any future editing via an authorization process, like restricting editing capacity to those with an authorized username and password. Another way is to provide a secure way to back-up the data collected. One practical way to do this would be to employ the services of a secure cloud back-up service, like IDrive Cloud Backup, which is a cloud service that periodically saves your data to its servers while you’re working (Kulkarni, R., 2022). Another way to ensure the integrity of the data would be to develop a policy that limits data entry and editing in specific ways. For example, columns from RfG data sets that deal with Seattle locations should not be allowed to be edited, unless the source proving the location, like the city government of Seattle, changes some protocol or the labeling for their locations, as a building or street could not possibly change locations under normal circumstances. Another way to ensure data source integrity would be to keep it secure from hackers or other businesses. One way to do this would be to use a virtual private network, like NordVPN, which would protect RfG’s data from any unauthorized users (NordVPN, 2022).

Most of the limitations to my data processing steps come from the conditions and restrictions set on my secondary sources. These include the inability to manipulate the data sets, and in some cases, see individual data points from the original data sets, as well as a required purchase before accessing the sources. An example of this is the secondary source *Statista – Share of joggers who went long distance jogging or running in 2021, by generation*, which requires a subscription and, while there is an overview of the methodology that Statista used to collect the data, the data set is not available to an outside user (Kunst, A., 2022). Other than this, any primary and secondary sources should also be cited if they were gathered by an outside organization.

### Data Analysis

• Analysis Concepts

* Some data analysis concepts that will be incorporated into the analysis process will be the *Hypothetico-Deductive Method*, the steps involved in data processing, as was discussed in the *Data Processing Steps* section, and *ratio* scaled data types. The Hypothetico-Deductive Method is a method of analysis based on the scientific method, where hypotheses are generated and systematically tested through measurements (Bougie, R. & Sekaran, U., 2020). In terms of what these means for RfG’s current business problems, hypotheses based on the research question will be tested against the data sources that were gathered for this report. Ratio scaled data is data on a continuous scale (Bougie, R. & Sekaran, U., 2020). An example of this are the ages present in *source 8*, “Statista – Share of joggers who went long distance jogging or running in 2021, by generation”. There are also other forms of data present in the data sources outlined. However, for the purposes of RfG’s business problem’s analysis, various forms of ratio scaled data are the most relevant. Other data analysis concepts that will be incorporated into this analysis have to do with data processing, which is project specific and has been discussed in the *Data Processing Steps* section.

• Recommended Solutions

* In terms of recommended analysis solutions to RfG’s business problem, I will first outline the software tool I will use. The software tool that will be used is the *SPSS data management and analysis program*, which can perform multiple forms of statistical data analysis and generate visual representations of any findings (Bougie, R. & Sekaran, U., 2020). This will be useful for our purposes, as it will help us analyze diverse sources of data, as well as display easily understood results. In terms of our analysis strategy, we can begin by separating our data sources into three categories for the sake of simplification, as well as to prepare for some of our data processing steps, such as harmonization. These categories will be finance related data sources, location related data sources, and marketing related data sources. Then, the data processing steps outlined earlier in the report will take place. After this, the sources will be analyzed to test the hypotheses based on our business questions, and the results will be reported, displayed, and discussed.

### Concluding Remarks

* Our consulting team’s main goal in this analysis process will be to provide useful information that will help Running for Glory’s management take actions to tackle their current business needs. As such, we remain ready to receive any feedback, new direction, and questions that RfG management may have as this process continues.
* **References**

Comber, S. (2021). Retail research in the era of Big Data: guiding the search for answers. *University of Liverpool.* <https://eds-p-ebscohost-com.ezproxy.snhu.edu/eds/detail/detail?vid=2&sid=a6c3ee8e-9253-420d-8fe6-0647b9eb60a1%40redis&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3d%3d#db=edsble&AN=edsble.839476>

Zhang, L. (2021). Proximity expenditure weights for Consumer Price Index: Audit sampling inference for big-data statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* <https://eds-p-ebscohost-com.ezproxy.snhu.edu/eds/detail/detail?vid=20&sid=494b52b4-7449-494e-b77d-dc356c3ddfa8%40redis&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3d%3d#db=bsu&AN=150144227>

Seattle Open Data. (2022). SPD Crime Data: 2018. *Seattle Open Data.* <https://data.seattle.gov/Public-Safety/SPD-Crime-Data-2008-Present/tazs-3rd5>

Seattle Open Data. (2022). SAIDI Electric Reliability Data 2014-2017. *Seattle Open Data.* <https://data.seattle.gov/City-Business/SAIDI-Electric-Reliability-Data-2014-2017/ahy7-fn62>

Bougie, R. & Sekaran, U. (2020). Research Methods for Businesses: A Skill Building Approach: 8th Edition. *MBS Direct.* [https://mbsdirect.vitalsource.com/reader/books/9781119561248/epubcfi/6/16[%3Bvnd.vst.idref%3DAc02]!/4/2/16/4[head-2-26]/1:5[MMA%2CRY](https://mbsdirect.vitalsource.com/reader/books/9781119561248/epubcfi/6/16%5b%3Bvnd.vst.idref%3DAc02%5d!/4/2/16/4%5bhead-2-26%5d/1:5%5bMMA%2CRY)]

Kulkarni, R. (2022). IDrive Cloud Backup: A Note from IDrive’s CEO, Raghu Kulkarni. *IDrive.* <https://www.idrive.com/cloud-backup>

NordVPN. (2022). NordVPN Official Website – Homepage. *NordVPN.* <https://nordvpn.com/country/usa/?vpn=brand&msclkid=618abd6e9bff1946cf15dff9f7ebd6e2&utm_source=bing&utm_medium=cpc&utm_campaign=Search%20-%20USA%20-%20Bing%20-%20Brand%20-%20En%20-%20Cpc&utm_term=nordvpn&utm_content=USA%20-%20Brand%7CExact>

Kunst, A. (2022). Statista – Share of joggers who went long distance jogging or running in 2021, by generation. *Statista.* <https://www.statista.com/statistics/227423/number-of-joggers-and-runners-usa/>