**Nyander Wheeler 08/26/2022**

**Predict the play for “Cash for Life” lottery**

**Executive Summary**

*I was hired by an Applications venture company, Illa Venture. Where there mission is, find the next new applications. Once they test well, they make the application and bring it to the public. Their team does intensive testing and decide if the idea will work, marketing wise through surveys. If the idea pass the test, the marketing team will distribute an impressive campaign.*

*For this application they want to make an app that can spill out numbers to users for lottery plays that is based on data. Because of that, they want me to write the codes and test the ability to predict the numbers based on numbers selected in the pass. I will also look at patterns over the years and present it to stakeholders.*

**Business Objectives**

*The goal:*

* *Can we provide some winning numbers 20%- 50% of the time and even get all of the numbers some of the time.*
* *If this is achievable can this be bundled and become an application.*

**Background**

*Apps that provide a consistence service to users are really the ones that are frequented often. While some people would not describe playing lottery as a valuable service, I would say most gaming apps do not provide a valuable service but users frequent the application daily. Keeping that in mind is where their mission came to light.*

**Scope**

*This project is used to:*

* *Decide should the application go to the next phase*
* *To find out if we can make predictions and be right with the evaluation questions for the next couple weeks.*

*This project is NOT used to:*

* *I do not have any at the moment*

**Functional requirements**

*For this project I will use:*

* *Tableau, Power point, Python, R and excel*

*Analysis:*

*Online I will find all the numbers played from the inception of the game, “Cash for Life” lottery. Once I make the data into an excel file, I will test it, but the Data wrangling will require some time because seven columns of info transfers into 1 cell. Once this is done I can start using the excel file in Python or R to build my analysis.*

* *My analysis must first answer questions 3 – 8. These questions look at the frequency of the data, such:*

1. *Are there 5 numbers that have played together more than once for each new draw?*
2. *Are there 4 and 3 that have played together more than once for each new draw?*
3. *What CB number is the most popular/was played the most?*
4. *What number was played the most out of the 5 numbers?*
5. *What 5 numbers were played the most on Saturday and Sunday?*
6. *What 6 numbers were played the most including CB on Saturday and Sunday?*

* *Question 1, one of the main questions, looks at all the data to determine, can it help me predict what three numbers (or 50% of the numbers) will play correctly 100% of the time during the 10- or 14-days period? We will look at the best statistical methods for this question. Based on my statistical flow chart, it looks like Stepwise Multinomial Logistics Regression is the way to answer this question.*
* *Question 2, the other main question, looks at all the data to determine, can it help me predict what six numbers (the total) will play correctly 50% of the time, during the 10- or 14- days period? We will look at the best statistical methods for this question. Based on my statistical flow chart, it looks like Stepwise Multinomial Logistics Regression is the targeted method to answer this question.*

**Personnel requirements**

*A data scientist or data and analyst can take on this project. Preferrable myself, Nyander Wheeler.*

*Must have knowledge of:*

* *Data wrangling*
* *Excel*
* *Analytical skills*
* *Statistical knowledge*
* *Machine learning*
* *Data Visualization*

**Delivery schedule**

**Data Retrieval**

* *Data wrangling: Pull the data from online fix it into excel, make the columns divisions: 2 days.*

**Process**

* *Planning the analytical process and guidance: 2 days*
* *Complete the analytical process and collect the results from the analysis: 1 to 2 weeks*

**Verification**

* *Find out if the analysis is correct: 3 days*

**Presentation**

* *Make data visualization boards to present information: 3 days*

**Implement the changes**

**Measure the data from the changes to confirm it is working**

**Assumptions**

* *Data is available*
* *Programming language works smoothly*
* Receiving timely feedback from your boss and other offices.

**Limitations**

*Internal to your project*

* *Time it takes to complete the project*
* *Personnel is healthy.*
* *The scope is not missing parts.*
* *Technical details are not available when needed.*

**Risks**

*External to your company*

* *Hurricane season or tropical storm*
* *Can we get the right data from offices or website to predict the economy of that area.*
* *Loss of power*
* *Covid – 19*