**Project 1 Team 7**

**Project Title:** World Happiness Project

**Team Members:** Sinan Cengiz, Ricardo Tirado, Kan Ming, Tiffany Kao

**Project Description/Outline:** We have a data set from 2015 to 2017 which contains happiness scores, GDP per Capital, life expectancy, generosity, government corruption trust score etc. of countries.We will examine the data set and answer following questions than put our finding in our final presentation to make it more visually understandable for consumer to help their decision-making process.

**Research Questions to Answer:**

**Question 1 -** Which region of the world has higher mean happiness score then other regions?

**Question 2 -** Is there any correlation between GDP per capita and happiness score?

**Question 3 -** Does countries has higher level freedom, also has higher happiness score? Relationship between freedom level and happiness level?

**Question 4 –** Does people have higher happiness level, also has higher life expectancy?

**Question 5 –** Does being more happy make people to have more babies or not. If there a correlation between happiness and fertility rate what is the direction?

**Data Sets to be Used:** Happiness rank and scores by country from 2015 to 2017. It is accessible from the link “https://www.kaggle.com/unsdsn/world-happiness”

**Rough Breakdown of Tasks:**

**Data scientist Sinan Cengiz :** this role would be primarily of someone who can work on large datasets (usually on Hadoop/Spark) with machine learning algorithms, develop predictive models, understand the “theory” – maths and stats behind the models and can interpret and explain model behavior in jargon-free language.

**Data Solutions Architect Ricardo Tirado :** A data engineer with a range of deep experience across several distributed technologies, and who also has good understanding of service-oriented architecture concepts and web applications.

**Data Engineer / Data Software Developer Kan Ming :** this role is for someone who has good understanding of distributed programming, including infrastructure and architecture. Typically, this person is comfortable with installation of distributed programming frameworks like Hadoop MapReduce/Spark clusters, can code in more than one programming languages like Scala/Python/Java, and knows Unix scripting and SQL.

**Data Platform Administrator Tiffany Kao :** A data engineer who has extensive experience across distributed technologies, especially managing clusters including production environments and good knowledge of cloud computing architectures (public clouds like AWS if using public cloud or OpenStack and Linux sysadmin experience if using private/hybrid clouds)