**Project Proposal and Requirements**

**Executive Summary**“Tripes & Travel.com” company is interested in more efficient marketing   
strategy in the coming year since they are introducing new vacation   
packages.

Utilized their previous year data on customers’ information and marketing   
details, our analyses will help the company to focus their resources more on   
marketing channels that yield the most result and eliminate or improve   
those channels that are performing poorly.

We will also help to narrow down on the segments of potential and current   
clients who are most interested in our packages in order to develop new   
vacation packages.

Independent variables or predictors:  
Customer’s data: age, occupation, gender, marital status, passport, own car,   
monthly income, designation  
Marketing’s data: type of contact, duration of pitch, product pitched, number  
of follow ups, pitch satisfaction score,   
Vacation packages: city tier, preferred property star

Dependent variable or variable we are predicting:  
number of persons visiting, number of children visiting, number of trips, pitch  
satisfaction score

**Business Objectives**In order to invest more efficiently in our marketing effort, we are going to   
determine the following with the Travel dataset provided:  
Which group of customers, marketing combination and vacation packages   
have the highest number of visits?  
Which group of customers, marketing combination and vacation packages   
have little to no influence on number of visits?

**Background**

“Travel.com” wants to reduce marketing cost while expand customer base   
by introducing new vacation packages. The Wellness Tourism Package will   
allow traveler to maintain, enhance or kick-start a healthy lifestyle and   
support increase sense of well-being.

Previously, 18% of customer purchased the packages with random contact,   
which cause high marketing cost.

The company’s goal is to improve on marketing for current market share as   
well as find out the best potential target customer base for the new Wellness  
Tourism Package.

**Scope**

As the group work on extracting, manipulating, pre-processing and generating predictions out of this data from Trips & Travel.Com dataset using statistical tools and programming languages such as R, Python, Excel, ggplot2, matplotlib and more. The use of a vast array of tools is necessary to carry out our data operations and will help the group gain insights about the travel company's products and processes.

**Functional requirements**

**Data Wrangling**:

The downloaded dataset should be successfully cleaned up for analyzing. The team should have a good understanding of what each column means, and how the values are measured.

Initial exam shows that there is no need to remove columns. There are NA value that will need to be remove. However, there will be sub-setting the dataset for analysis.

The datatypes for each column should also be converted to a usable format for the needed analysis.

**Data Analysis**:

R, Python seems to be sufficient with what we are planning to do with this dataset. However, others analysis program might be added as needed.

Our analysis will exam the most efficient marketing target and method that yield the most result with the lowest cost for company new travel package. In order to do that, we will exam the relationship on factors in dataset on number of trips made and numbers of visitors per trip. Expected analysis to perform including anywhere from linear regression to ANOVA.

**Data Visualization**:

Findings with be presented in a Power Point slideshow using visuals from program such as ggplot, Tableau, matplotlib and classic Excel graphs.

**Presentation**:

Final presentation will be at a prior schedule time via Zoom by all members of the group.

**Personnel requirements:**

Marissa, Alex and Quy will be working closely on this project together. Each team member will take turns being the scrum master for a week.

Daily communication would be Slack to problem-solve or to check in on work progresses. In addition to that, there is a weekly meeting to review the past week workload and plan out their next sprint.

Beside team meetings, there is also a weekly meeting with instructor Daniel to report progress, seek assistant and/or advise for the direction the team is taking on Wednesdays

**Delivery schedule**

Week 1: Import dataset into preferred software to begin data wrangling. Any unnecessary data should be removed. Set up Github and Kanban board.

Week 2: Study the dataset and ask questions. What are some possible correlations? Is the data normally distributed? What are some predictive models we can make from it? Visualize the data to see if there is any interesting findings.

Week 3: Modeling/Optimization (Combined Stepwise - Forward and Backward Selection) and Machine Learning (Random Forest.)

Week 4: Review and validate findings from the previous week, and draw insights/conclusions.

Week 5: Compile findings into a Power Point slideshow. Go over it with their instructor and friend/family member to ensure that the presentation is clear and logical. Work on the style and layout of the presentation so it is delightful on the eyes.

Week 6: Make final touches to the Power Point presentation. The group should not attempt to come up with a brand-new analysis at this time. They should practice presenting at least a couple times with the group, and at least once with their instructor.

**Other requirements:**

**Budget:**  all require programs and data should be free of charge.

**Internet**: We all rely on internet to get things accomplished. In case of an outage,   
teammates can use a phones hot-spot, public wi-fi, or even a friend’s house is it is available.

**Equipment**: Each team member should have a computer or laptop with all require software installed and function well.

**Environment**: Each team member should have a working environment that enable them to focus on the projects with minimal distraction.

**Assumptions:**

All listed programs either installed or online should run smoothly.

The data set will provide us with needed answers and results.

Computers and internet resources should also be a safe assumption, that they will work.

All outside influences remain the same. No further restrictions due to Covid or unexpected natural disasters that might hinder, delay or prevent the team to meet deadline.

**Limitations:**

**Time-zone**:

The different time zone cause limitation to duration of meeting due to other time sensitive normal life activities.

**Expertise:**

Team members are not experience data scientist. Therefore, there are an element of lacking of experience in data science related knowledge and experience.

Team members are not familiar nor connected to the company prior to this project. As a result, the knowledge regarding this project limits to what provided in the dataset.

**Risks:**

**Data loss:** to prevent losing data permanently make sure it is backed up and submitted via   
Slack or The Cloud or another type of file uploading system. Another risk is mental block,   
ensure to utilize each other, slack, our teacher, or online resources to continue the project.

**Power**: Although laptop has a battery, team members should plan for alternative working location in case of power outage.