**Team Entrepreneur-sheep**

**Business Analytics Project Brief**

**Project NYC Yellow Taxi**

Date: May 11, 2016

Version: 3.0

Authors: Brenda (Yuyao) Huang (yh1862)

Brenda Truong (bst259)

Christoph Krinner (ckk264)

Ginda Bastari (gb1689)

Jungmin Park (jp4456)

Course: MG 9753 Business Analytics

TABLE OF CONTENTS

[Project Summary](#_u19enqszf9rf)

[Overview of Needs](#_pdzt374404it)

[Project Deliverables](#_f2f97vmgg54m)

[Data Summary](#_clck2bqkj9kr)

[Proposed Methodology](#_jia29xn3bt6h)

[Project Success](#_jrxrer7lg1j3)

[Personas and User-Stories](#_dcmhu6v6rff9)

[Project Plan and Milestones](#_5566kogrb19s)

[Project Roles and Responsibilities](#_yn9sdgo066qx)

[References](#_38l98uusrtf5)

PROJECT NYC YELLOW TAXI

# Project Summary

Taxicab service plays a vital role in public transportation in NYC. Yellow taxicab drivers earn their income by cruising the road network looking for a passenger alongside the street. This service is often inefficient. We will explore and analyze data to model the potential profitability of locations given a current location, census data, weather data and time of a taxi driver.

We will use information from Public datasets containing diverse variables. We will use linear regression to find out the relationship between variables and performance.

This project will help us to make personalized suggestions to a taxi driver based on location, time, weather conditions and demographic information. Our research and data analysis of diverse routes in the NYC Yellow Taxi dataset will be a great resource for our user case.

# 

# Overview of Needs

This project aims to analyze publicly available data and derive predictions about what factors foster profitability success for yellow taxicab drivers in NYC.

# Project Deliverables

The following are the proposed deliverables for this project:

1. Data exploration and modeling scripts: The team believes in open data. We will release our exploration and modeling documents.

2. Project brief: By using project brief, we can make a plan and organize what we need to do. We are putting together the architecture and design of our project from a high level and project management perspective in the project brief.

The following are “reach”, or “nice-to-have” deliverables for this project:

1. Datasets : Databases and datasets collected.

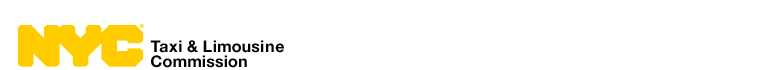
2. Data Visualization files: We will use Tableau to present the processed information visually.

2. The final project slides: We will present the final slides showing the outcomes generated.

# 

# Data Summary

We source all our data from New York City Taxi and Limousine Commission (T&LC), United States Census Bureau and National Centers for environmental information. Data includes:



* Pick-up Location
  + Longitude, Latitude
* Pick-up Time
* Total Fare Amount
* Tip Amount
* Weather Condition such as wind speed, precipitation, maximum temperature and minimum temperature.
* Demographic condition such as number of population, percentage of white, asian, etc.

Out of Scope: The project only analyze data from May 2015 due to time limitations.

# Proposed Methodology

To identify the right deliverables and ensure that our analytics process results in successful business outcomes, we will use the CONVO methodology.

● Context

* Cab drivers will benefit from our analysis by knowing where and when to drive and other techniques to make more fares.
* Taxi Company will be interested in the results since they will know where to allocate the cabs and resources.

● Need

* How is every trip coordinated to local census tracts, neighborhoods and other external factors?
* What are the stories and meaning from the data?

● Vision

* Underlying coordinates.
* Powerful Predictors.
* Story Telling.

● Outcomes

* Regain competitiveness against Uber and other car services in the market.

# Project Success

Our project will be successfully completed if we can find three to four factors that significantly predict yellow taxicab drivers success and successfully train our predictive model.

# Personas and User-Stories

Personas and user stories were defined as a guideline to help us set the requirements in outlining the proposed solution. As noted by Goodwin, using personas as archetypes will bring us the ability of using people rather than abstract ideas as presentation (Goodwin K., 2009). Thus, they will help us to see the archetype in an intellectual and emotional way. We introduce our personas:

* Matt is a NYC Taxi driver, he works part time in the NYC 5 boroughs: Manhattan, Queens, Brooklyn, Bronx and Staten Island. He got his taxi driver licence from the New York City Taxi and Limousine Commission (T&LC) . He has been working only on weekdays. As a owner of a yellow taxi in NYC he wonders what are the factors that affect his profits? She also wonders how can he get more profits in the month of May. He tried to find reliable data in diverse websites, but there is always a subscription fee that impede the full data analysis.

The personas serve as the foundation to build user stories. As Cohn explained, user story describes the functionality that will be valuable to potential users and will help the designer to formulate and present the requirements from the user perspective. Cohn also proposed a formula to create a user story: As (role) I want (something) so that (benefit). (Cohn M., 2004)

The below user stories were created based on the persona’s identity, goal, and possible benefit when using the proposed solution:

* As a NYC Yellow taxi driver I want to know what are the factors that influence my profits so that I can strategically earn more.

# Project Plan and Milestones

We present the project task plan with phase completion dates and key milestones.

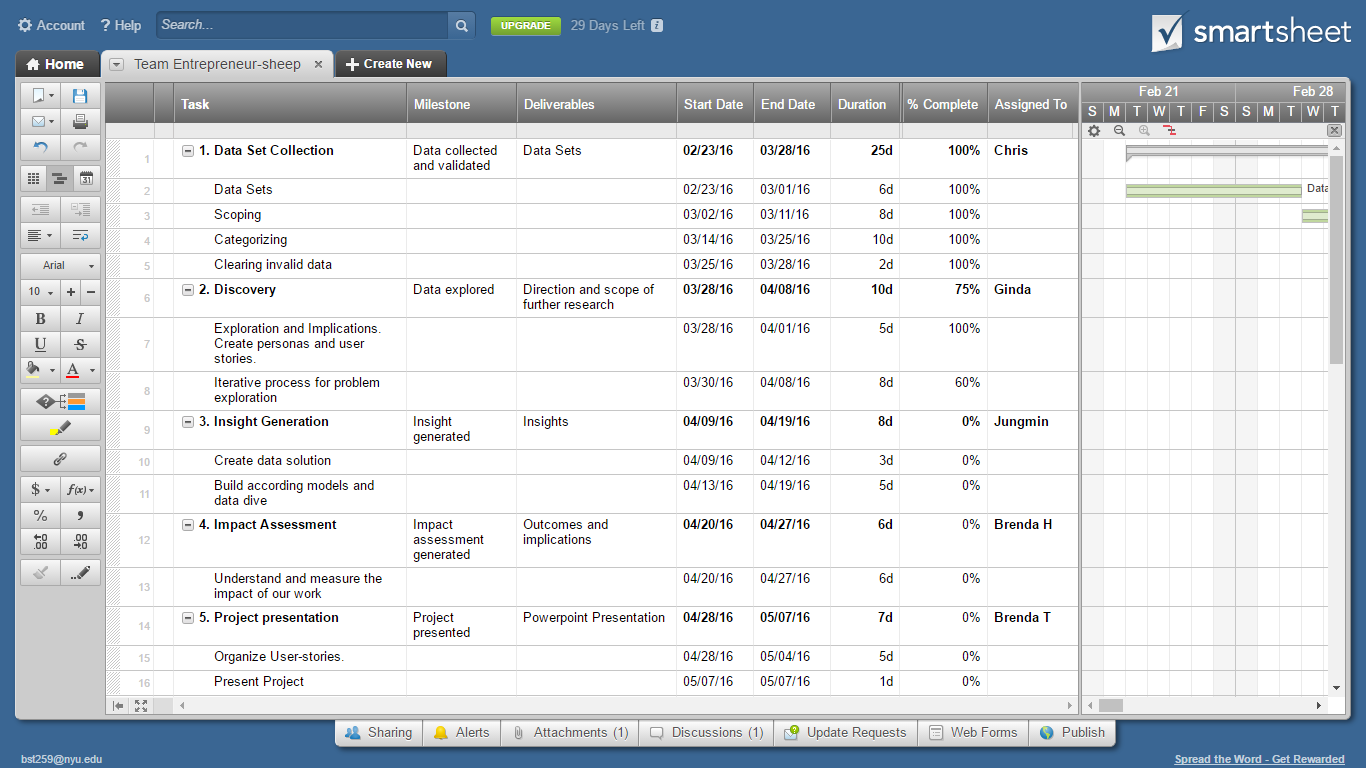


Fig. 1a - Gantt Chart that illustrates summary elements of our project.

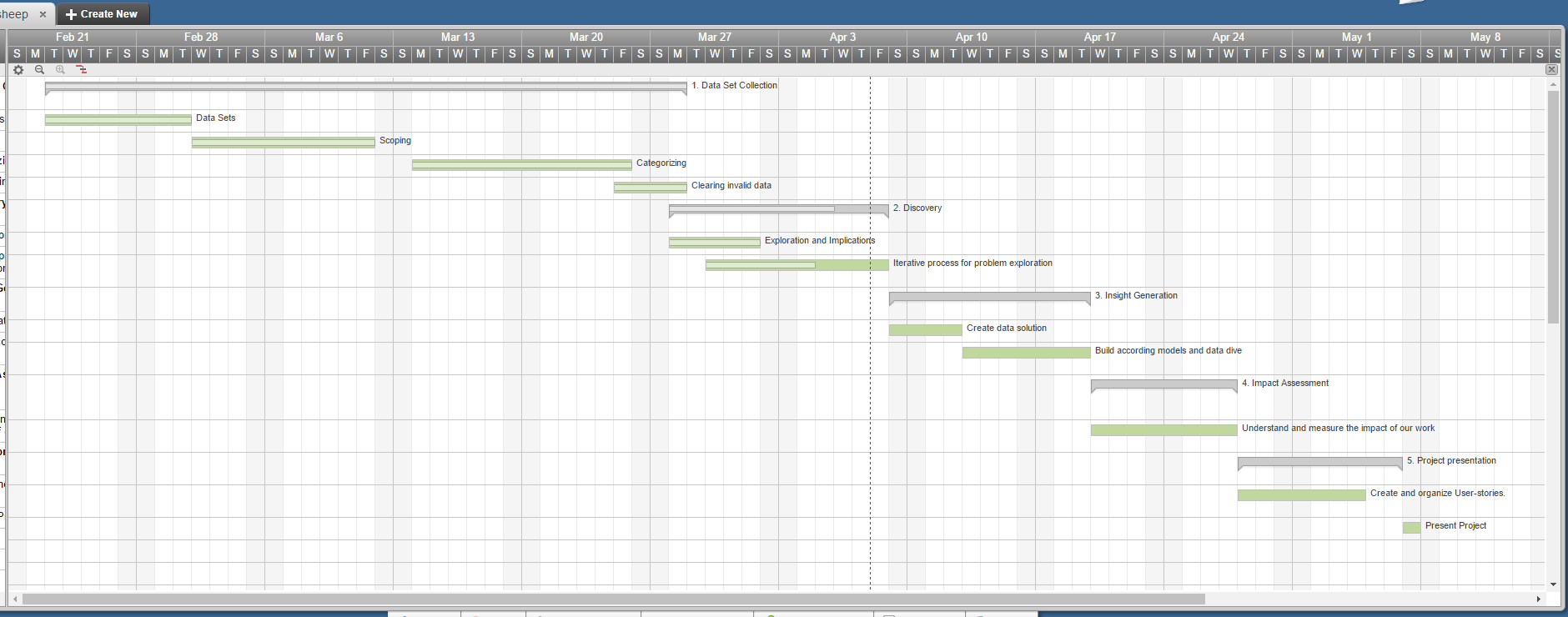


Fig. 1b - Gantt Chart that shows the project schedule.

# Project Roles and Responsibilities

|  |  |  |
| --- | --- | --- |
| Name | Role | Responsibilities |
| Brenda Truong (bst259) | Technology Coordinator | Data Visualization, Which tools to use, how to present data |
| Brenda (Yuyao) Huang (yh1862) | Consolidator and Editor | -Set up project plans, strategies and bigger pictures  - Consolidate the research and result together  - Edit and write up reports and research deliverables |
| Christoph Krinner (ckk264) | Data Analyser | - Explore data sets  - Compute calculations in R and Tableau |
| Ginda Bastari (gb1689) | Data Diver | - Discover relevant data  - Exploring data sets |
| Jungmin Park (jp4456) | Team Coordinator | - Set Weekly targets. - Set timelines for every team assignment. - Plan the location and time of the weekly meetings. - Set timelines for every team assignment |

# References

Clark, W. (1952). *The Gantt chart; a working tool of management*. London: Pitman.

Cohn, M. (2004). *User stories applied: For agile software development*. Boston: Addison-Wesley.

Goodwin, K. (2009). *Designing for the digital age: How to create human-centered products and services*. Indianapolis, IN: Wiley Pub.