Final Project Rubric

# MSDS 596 - Group 6

Fall 2021

One of the deliverables students are required to submit for the successful completion of MSDS 596 is a final project. The final project will consist of an oral presentation. You will lead the audience through a set of visualizations you have developed, telling a compelling story to your audience.

# Group Work

Students are required to work in a group of four or five students. Project groups were set in Week 2. Please navigate on Canvas to "People" -> "Project Groups" to see which group you belong to.

# Data

Students are required to select their own data set. Students are encouraged to select data on a subject about which they are passionate. Students should meet with the instructor at least once during the duration of the course to review the data set they have selected.

Here is the link where you can find the most recent Kickstarter data. <https://webrobots.io/kickstarter-datasets/>

# Deliverable

Each group must submit the following to the instructor to receive credit for the project.

## Submitted as a group

1. All data—or a link to all of the data—employed in the generation of analysis/visuals
2. A set of 8 to 10 static visualizations (2 per student in the group), collected in slide presentation format (PowerPoint, Keynote, Google Slides). This should be submitted as a pdf. All visualizations should be high-resolution images, not grainy nor blocky. Note in the corner of each slide which student "owns" the slide (i.e. was responsible for the analysis and visualization).

## Submitted individually

1. All code used to wrangle, scrape, clean, process, aggregate, and visualize data.

* All code used for the above-mentioned tasks will be written in Python 3 using the pandas package on Google CoLab. Provide a link to the notebook used for the code.

# Presentation

On the final day of classes, student groups will present their work. Groups will have a total of 15 minutes to orally present their entire work, including introducing their data set, presenting all visuals, and concluding the presentation. Within the 15 minute timeframe, the last 2-3 minutes should be reserved for questions.

# Grading

### QUALITY OF ANALYSIS (INDIVIDUAL) [25%]

* sufficiently rich and complex questions are addressed via the data
* analysis is performed accurately, exploring possible reasons for interesting findings

### CODE CLARITY AND QUALITY (INDIVIDUAL) [25%]

* code is cleanly written
* code is commented where appropriate
* markdown is used to break up sections of code and/or explain code

### QUALITY OF VISUALIZATIONS (INDIVIDUAL) [25%]

* visualizations are aesthetically pleasing
* visualizations support the theme of the presentation
* there are a variety of visualizations employed, i.e., presentations should not consist of variations of the same graph or type of graph repeatedly
* use of and adherence to design principles outlined in the course

### GROUP PRESENTATION & STORYTELLING (GROUP) [25%]

* ability for a group to communicate their message via the oral medium in a clear fashion
* the story/impetus for the use of data is compelling
* a singular idea/objective is woven through the presentation/visualizations
* use of and adherence to data storytelling techniques covered in class

# Things to Do:

Come up with 3 questions that you can ask about the data? (possibly related as you want to tell the story with your slides)

[Police Department Incident Reports: Historical 2003 to May 2018](https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-Historical-2003/tmnf-yvry)

**LINK TO FINAL PROJECT FOLDER**

https://drive.google.com/drive/folders/15eeK4NhcWYtakHeIEAPEIN7g2E-H94Sl?usp=sharing

To combine all the kickstarter files into one dataframe:

# load libraries

import pandas as pd

import numpy as np

import glob

# import .csv files from data folder on Google Drive

path = r'/content/drive/MyDrive/2021\_Fall\_USF\_PMS\_Biotechnology/MSDS\_596/final\_project/data/' # use your path

all\_files = glob.glob(path + "/\*.csv")

li = []

for filename in all\_files:

df = pd.read\_csv(filename, index\_col=None, header=0)

li.append(df)

ks\_df = pd.concat(li, axis=0, ignore\_index=True)

ks\_df.head(5)

# Leeza Sergeeva

415-378-4688

[esergeeva@usfca.edu](mailto:esergeeva@usfca.edu)

On campus Monday (6:30pm-9pm), Wednesday (4:30pm - 9pm), can come Thursday/Friday after 6 pm.

# Ezana Kemmer

707-338-5869

[eakemmer@dons.usfca.edu](mailto:eakemmer@dons.usfca.edu)

These are just some questions that came to mind

* What % of crimes occur during the day vs at night?
  + Might be too complicated, but the dataset has a time column which we might be able to filter? .
* Which district had the most amount of incidents?
  + I feel like this one might be the easiest? All you’d have to do is group by district and total number of crime?
  + Similarly, maybe we can compare between districts for categorical data like which types of crimes are the most widespread in each district? Again, might be overcomplicating it, but just brainstorming.
* **Which day of the week has the most number of crimes?** 
  + Just finished this. Will construct a bar chart showing the difference.
* **Maybe which crime had the highest rate/largest number of incidents?** 
  + **Basically what are the most popular types of crime in SF?**
  + **Have this done**
* How has the total number of crimes changed over time?
* Which street had the most incidents?

**My Completed work**

1. Made a histogram and line chart showing the frequency of the top 5 most popular crimes over time
2. . Made a bar chart that categorized the total number of crimes by each day of the week.
   1. Also included another bar chart that segmented crime by time of day(night vs day).

# Fernanda Nunes

415-425-9277

[fvalledesallescoelho@dons.usfca.edu](mailto:fvalledesallescoelho@dons.usfca.edu)

1. 10 Worst and 10 best neighborhoods by # of incidents
2. # of Incidents by Neighborhoods

# Luis Rodriguez

Line Graph on increase/decrease fo crimes

Pie graph on percent of crime by the district

661-565-5768

[lrodriguez24@dons.usfca.edu](mailto:lrodriguez24@dons.usfca.edu)

Global Kickstarter

Success / Failure

Popular Categories

Pledge Amount average for successful projects

Deadline - launched\_at

<https://www.statista.com/statistics/222489/most-successful-completed-kickstarter-projects-by-total-funds-raised/>

<https://www.statista.com/statistics/235405/kickstarter-project-funding-success-rate/>

<https://www.statista.com/topics/2102/kickstarter/#topicHeader__wrapper>

<https://www.statista.com/statistics/310218/total-kickstarter-funding/>

<https://github.com/matplotlib/cheatsheets#cheatsheets>

Join Zoom Meeting

<https://zoom.us/j/7598076565?pwd=SHVLZ0JwOXlmMWEwRlIrdyt5eURwQT09>

Meeting ID: 759 807 6565

Passcode: TAhour

Collab Notebook - Leeza Sergeeva

<https://colab.research.google.com/drive/1K3mzNzV58gG5iw5PllcZJhzaQjS9EtRJ?usp=sharing>