ESOC 214 Final Project Proposal

Adriana Picoral

Fall 2020

TITLE What is the work-in-progress title for your project? (you can change this later)

1. INTRODUCTION

Introduce your project, you might want to address the following questions:

- What is the question that you are interested in?
- What is the application of your project?
- What larger issues/questions does your project address?
- What insight does your project provide?
- Why does your project matter from a big picture perspective?
- Does your project address a business use case, if so, what is it?
- Does your project inform societal issues, if so, how so?

2. RESEARCH QUESTION

Clearly state your research question(s). Make sure your question addresses more than one layer of analysis (i.e. include different aspects of how your question interacts with your data). Consider how one or two numeric variables change over time across different groups of people.

- For example, if I was interested in how temperature affects the number of bike rentals, my questions would be "Do higher temperatures correlate with a higher number of bike rentals? Is this correlation the same across different groups of riders (i.e. age, gender)?"

3. METHODS

3.1 DATA

Clearly state what your data set is, and what variables you are going to use to answer your research question.

What type of evidence is needed to answer your question? For example, if I was interested in how temperature affects the number of bike rentals, I would need to show that as temperature increases that the number of bike rentals also increases.

What data set are you going to use to answer your question? Do you know where to find your data set? Document where you acquired your data from.

3.2 DATA WRANGLING

What are the variables within the data set that allow you to answer your question? In the case of the bike question, I would need at least the following variables: date, number of rides, and temperature.

If you don't have the above variables in your data as columns, can you manipulate your data to create them? For the bike question, I only have individual ride data, not the total number of rides in a day. But, I know I can use a group_by to count up the total rides each day.

What issues do you expect to find in your data? (you might need to briefly explore it to answer this question) This could be data cleaning issues, missing values, etc. How do you plan to solve them?

3.2 DATA VISUALIZATION

What plots do you plan to display to visually show how variables relate to each other in a way that answers your question? In the bike example above, I would need a figure showing that the trendline of temperature closely mirrors the trendline of the number of bike rides.

4. OTHER CONSIDERATIONS

What's one issue you might don't know how to solve? Specifically, what is your input and what's the desired output? Do you have what you need for your input? Do you know how to generate the output?