**ADS-506 – Final Team Project Start Form**

Fill out this form and submit it by the end of Module 2 in Blackboard.

Team Number: 6

Team Leader/Representative:

Vanessa Salazar

Full Names of Team Members:

1. Michael Nguyen

2. Vanessa Salazar

3.

Title of Your Time Series Final Project:

Streets Repair Projects Reduction in San Diego

Motivation for choosing this project:

Street repairs are a constant in the daily lives of San Diego residents. These repairs include regular maintenance of roads but also more extensive road repairs which require entire streets to be excavated. By collecting information on road repair completion times this information can be provided to local residents to improve transparency on road repair duration. The information can also be used to predict how long road repairs take to be completed and therefore applied to determine future road work project completion times.

Problem Statement: Short Description of Your Time Series Project and Objective(s):

The City of San Diego is in need of completion times for road repair projects to forecast future timelines of street repairs. Road repairs not only apply to preventative maintenance of streets but also to large projects such as reconstruction of entire roadways. By understanding completion times and patterns of road repairs future road repairs can be made more efficient and transparent. Using previous data collected on City of San Diego road repair from July 2013 through December 2021 which includes types of street repair along with start, end, and completion dates. Road work crews can attempt to remain on schedule as much as possible and residents of San Diego can be aware of how long the inconvenience of road repairs will affect their daily lives.

Name of Your Selected Dataset: sd\_paving\_dataset\_v1

Description of your selected dataset:

The data set includes the pavement event ID, the segment ID in SAP, the Project ID, project title, project manager’s email and phone number, the scheduled start and end date of the project, when the street went into moratorium, the current status of the project, the type of work the last pavement project executed and the length and width of the project in feet.

Data source, number of variables, size of dataset, etc:

Our data is in a CSV file with 19 variables. The dataset is 6.06MB and contains 24,935 rows and 19 columns of data.

Notable findings from your initial EDA:

- Most of the status is post construction

- There are many TBD in resident\_engineer, showing that we do not know who did the work. This means there are some null values.

- The dates are formatted so that it is easy to use

Github link:

https://github.com/mycunguyen/Streets-Repair-Projects-Reduction-in-San-Diego