**Corridor trend related to walk-in businesses project proposal**

1. **Scope of the project-**

The Community Redevelopment Area (CRA) of the south St. Petersburg is working towards development of Commercial corridor Planning program aiming at funding infrastructure improvement and planning efforts as future public investment. With this project, the CRA is aiming to analyze and interpret the corridor trends in relation to the walk-in retail business like- shopping, entertainment, restaurants, etc. The goal of this project is to provide a statistical analysis of the walk-in businesses situated in each of the commercial corridors and to determine the corridor with highest number of walk-in businesses. This analysis would help CRA plan more strategically in regard to the investments to be made around the city of St. Petersburg by, as having exact percentage of the walk-in business in each of the corridors would help the CRA determine the investments they would have to make to make any infrastructure improvements towards these businesses.

1. **The Dataset-**

For this project, we will be using the “Business Tax ID Data 2015” dataset, which includes information like address, business name, industry, walk-in status, the eight corridors of interest, and a PIN that is unique to each business.

1. **Methodology-**

This project requires us to determine the percentage of walk-in business on each of the eight corridors for which we would be using statistical tools like- Excel and Python programming.

**Our plan of action to achieve this outcome**-

1. To segregate the **industry column** from the dataset, to filter out business falling under the category of **retail businesses**.
2. The **Walk-in column** in the dataset (target variable) is a categorical column, categorizing the businesses to be walk-ins or not. We will filter out all the walk-in rows for our analyses. Going through the initial exploration of the dataset, we noticed that this column has several missing values. The dataset has 2404 records out of which 1301 records are missing in this column. We could think of two ways to handle this situation.
3. Impute all the missing values with the most common attribute in the column, as it is a categorical column, but this would introduce bias.
4. Delete the rows with missing Walk-in fields.

We are awaiting a response to determine the best course of action for this column.

1. The dataset has several duplicate records, as there are certain businesses that appear to be more than one business. To identify the presence of one business in multiple industries, we will use the **column PIN**, as each business has a unique PIN.
2. We plan on using python libraries to calculate the percentage of walk-in businesses in each of the eight corridors (each corridor has a separate column in the dataset), and then sort these columns based on the number of walk-in retail businesses they hold to provide a comparative ranking of the columns with most walk-in retail businesses.

To present our findings we plan on using visualization tools like- **PowerBI, matplotlib library in python and excel.**

1. **Anticipated business outcomes-**

With this analysis, CRA would have stats available for the number of retail walk-in businesses they have on each of the eight corridors. This would help them in their Commercial Corridor Planning Program, as they could determine their investment strategies, by planning the amount of investment they would have to make towards the infrastructure improvement of these businesses.