**Analyze and compare corridor trends related to "walk-in” retail businesses in 2015**

**Group 6**

***Nitika Mishra***

***Sri Venkata Likhitha Duggi***

***Vandana Sirpa***

***Aswith Reddy Kovvuri***

***Manohar Reddy Pasham***

Table of Contents

[Introduction 3](#_Toc132893179)

[Methodology 3](#_Toc132893180)

[***Data Description*** 3](#_Toc132893181)

[**Approach Followed:** 3](#_Toc132893182)

[Results 4](#_Toc132893183)

[Conclusions 6](#_Toc132893184)

[Link to Visualizations 7](#_Toc132893185)

[Appendix 7](#_Toc132893186)

# Introduction

The South St. Petersburg Community Redevelopment Area (CRA) is in the process of creating a new Commercial corridor Planning program. This program aims at funding future public investments in planning initiatives and infrastructure improvements.

This project aims to provide a statistical analysis of the walk-in businesses in each of the eight corridors for ***Business Tax ID Data 2015***. Below are the research questions this project will focus on:

1. To find the percentage of walk-in retail businesses in each of the eight commercial corridors.
2. To find which corridor had the highest number of walk-in retail businesses.

This analysis would help CRA plan more strategically and make an informed decision regarding the investments around St. Petersburg.

# Methodology

## ***Data Description***

For this project, we used the ***Business Tax ID Data 2015*,**and the source of the data is - the City of St. Petersburg’s Business Tax ID Data for properties on commercial corridors within the South St. Petersburg CRA. This dataset consists of 21 columns and 2405 rows. Address, City, State, Postal Code, Control Number, Year, License Number, Business Name, Pin number, Business Status, Industry, Walk-in, CRA, and the eight different corridors are the attributes of this dataset. Our main focus was on the **Control number, PIN, Industry, Walk-in, and the eight corridor columns**.

## **Approach Followed:**

1. The business status variable in the dataset represents the active or inactive status of the businesses. There were 1032 records with inactive business status. These records were removed from the analysis, and the remaining 1372 records were considered.

2. The Control Number variable in the dataset is the unique id for businesses. Using the remove duplicate functionality in the data tab of Excel, 88 duplicate rows were found and removed. 1283 rows remain in the dataset, with only one row for each control number.

3. After duplicates were removed from the control number column, the next step was to deal with the empty rows in the Walk-in Column. This column represents if the business is a walk-in or not and is represented as "Yes" and "No," respectively. As the empty rows were to be considered towards the value "No," these rows were replaced with the value "No" using the find and replace function in Excel, and a total of 609 replacements were made.

4. The next step was to remove duplicates based on the PIN. Duplicates were removed from the PIN column only if they had the same Industry. For example, if the PIN "23 31 16 78409 001 0011" had five duplicate values, out of which four businesses are retail trade and one is from the food service industry, then three businesses within the retail Industry were removed, leaving two instances of this particular PIN in the dataset with one business in the retail trade industry and the other one in Food Services. Both were used in the analysis as they signify unique businesses.

We used the conditional duplicate removal functionality in the data tab of Excel to remove duplicate occurrences from the PIN column if those occurrences have the same Industry. 184 duplicate values were removed, which resulted in 1187 unique rows.

5. Then the excel file was imported into Python notebook to calculate percentage of walk-in retail businesses in each corridor and then do a comparative ranking of the corridors.

6. The cleaned data was then fed to Tableau for performing visualization to represent the percentage of walk-in businesses in each of the eight corridors and to rank them according to the number of walk-in businesses each corridor has.

A flow chart of the steps is available in Appendix B.

# Results

The first research question is to determine the percentage of walk-in retail businesses in all eight corridors. We visualized this in the form of a pie chart for every corridor for a clear representation. After the removal of any duplicate businesses, we calculated the percentage of walk-in retail businesses using the formula -

(Total number of walk-in retail businesses in a corridor/ Total number of businesses in that corridor) \* 100.

We used filters in Tableau to display pie charts. Among these corridors, ***Dr MLK Jr St Corridor has the highest percentage of walk-in retail businesses, with 66.67% of the businesses being walk-in retail businesses.*** 18Th Ave S Corridor has the second highest percentage of walk-in retail businesses, with 64.7% of the businesses being walk-in retail businesses. 34th St Corridor and 22nd St S Corridor have 47.9% and 44.44% walk-in retail businesses, respectively. 16th St Corridor and 49th St Corridor share the same percentage (42.86%) of walk-in retail businesses. Central Ave Corridor has 34.93% of its total business as walk-in retail businesses, and the lowest percentage of walk-in retail businesses was observed in 5th Ave Corridor, with 25% of the business as walk-in retail businesses. Pie chart representations for all the corridors are available in Appendix A.

The most interesting part of this analysis was that during the initial findings, the Central Ave Corridor stood out with 80 walk-in retail businesses, which was the highest number of walk-in retail businesses observed when compared to the other corridors, but the overall percentage of walk-in retail business was low for Central Ave Corridor in comparison to other corridors.

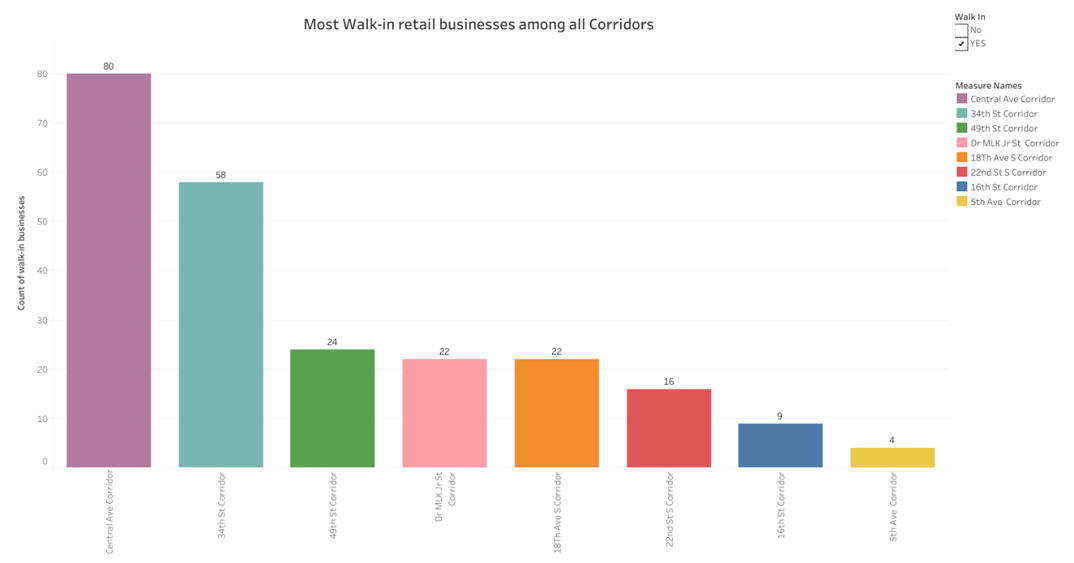
Below is the pie chart representation for the same.

*Chart, pie chart

Description automatically generated*

The second research question was to provide a comparative ranking of these corridors to figure out which corridor has the highest number of walk-in retail businesses. ***Central Ave Corridor* was leading with 80 walk-in retail businesses**, followed by 34th St Corridor with 58 walk-in retail businesses. Next in the line was 49th St *Corridor* with a total of 24 walk-in retails, followed by MLK Jr St Corridor with 22 walk-in retail businesses. 22nd St S Corridor has 16 walk-in retails; 16th St Corridor has 9, and the lowest number of walk-in retail businesses was observed in the 5th Ave Corridor with only 5 businesses.

Below is the graphical representation of the comparative ranking of the number of walk-in retail businesses in the eight columns:

**

# Conclusions

The analysis and comparison of corridor trends related to walk-in retail businesses have revealed important insights about walk-in retail businesses in each of the corridors. After careful examination of the data and manipulating it to remove duplicate and inactive businesses, we successfully determined the ***percentage of active walk-in retail businesses*** in each of the eight corridors. We inferred a ***comparative ranking among these corridors*** for the number of active walk-in retail businesses.

With this analysis, CRA and the sponsors would have precise stats for the number of retail walk-in businesses on each of the eight corridors. This would help them determine their investment strategies by planning the investment they would have to make toward the infrastructure improvement of these businesses.

Comparing the businesses with the highest number of walk-in retail stores in each of the corridors and their performance comparison in every corridor would provide opportunities to analyze this data in these aspects further, providing a more comprehensive understanding of the trends in walk-in retail businesses. This would help develop more targeted and effective strategies to make future public investments.

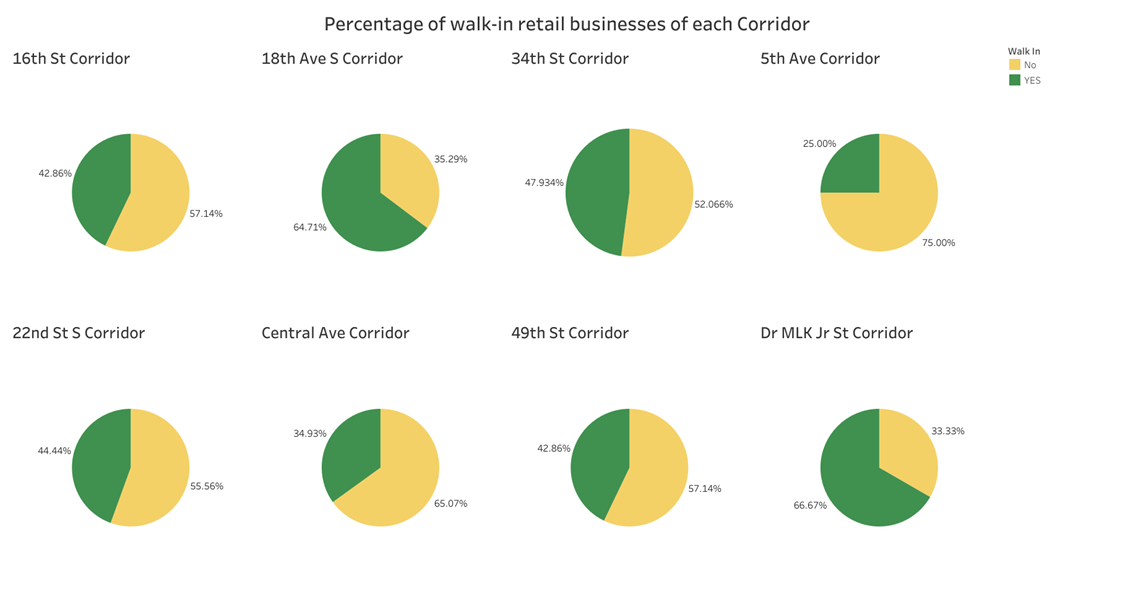
# Link to Visualizations

<https://public.tableau.com/app/profile/vandana.sirpa/viz/EIS_Team6_ResearchProblem_1/percentage?publish=yes>

<https://public.tableau.com/app/profile/vandana.sirpa/viz/EIS_Team6_ResearchProblem_2/count?publish=yes>

# Appendix

***Appendix A: Percentage of walk-in retail businesses in each corridor***



***Appendix B: Methodology***

Diagram

Description automatically generated