**Course Project User Guide**

# Introduction

The City of St. Petersburg is a vibrant and diverse metropolitan area with a thriving commercial sector. As part of its ongoing efforts to promote economic growth and development, we have embarked on an experiential learning project to analyze the industry mix and diversity across each of its eight commercial corridors. The project aims to provide valuable insights into the economic landscape of the city, which can aid in strategic planning activities related to investments in and around these areas.

Research Question for the project: Exploring the Industry Mix and Diversity in Each Commercial Corridor from business tax ID data 2023.

The project's central focus is to explore the industry mix and diversity of the commercial corridors in St. Petersburg. We have conducted an individual analysis of each of the eight commercial corridors in St. Petersburg to determine the range of industries present and their respective levels of diversity.

# Methodology

**Description of the Data:** This dataset contains information on businesses located in the City of St. Petersburg, including their addresses, city, state, and postal code.

Each business is assigned a unique "Control Number" that remains the same every year and a unique "PIN #" that identifies the property. The "Business Status" column indicates whether the business is currently active ("A") or inactive ("I").

The "Industry" column lists the primary industry of each business, but some businesses may operate in multiple industries and, therefore, could be duplicate rows in the dataset. The "Walk-In" column identifies whether a business relies on pedestrian foot traffic to thrive and contributes to a pedestrian-oriented retail/services shopping environment.

Additionally, the dataset includes information on the location of each property within specific commercial corridors, including 16th St Corridor, 18th Ave Corridor, 22nd Street Corridor, 34th Street Corridor, 49th St Corridor, 5th Ave Corridor, Central Ave Corridor, and DR MLK JR St Corridor. Each of these columns can be used to filter data for businesses located on or within a specific commercial corridor.

Finally, the "CRA" column indicates whether the property is located within the Community Redevelopment Area (CRA), and the "Neighborhood" column lists the name of the neighborhood where the property is located.

**Methodologies Used:** After reviewing the dataset and project requirements, we conducted descriptive data analysis following the CRISP-DM process model. The following steps were taken:

1. **Business Understanding:** The goal was to analyze the industry mix/diversity across each of the eight commercial corridors of the City of St. Petersburg from historical data. The research question we worked on was "What is the industry mix/diversity on each commercial corridor?" We defined the project's scope as analyzing data from each commercial corridor, which included data on businesses' address, city, state, postal code, control number, PIN #, business status, industry, walk-in status, CRA, neighborhood, and corridor.
2. **Data Understanding:** we examined the available data and identified several constraints that could affect our analysis. For example, we found that certain businesses appear to be more than one business but are registered as separate businesses in the dataset. We also found that some single businesses target more than one industry and have multiple rows. We identified these constraints to prepare and preprocess the data before conducting descriptive data analysis.
3. **Data preparation and preprocessing:** One of the key tasks in this step is removing duplicates from the dataset based on certain constraints. For example, a massage parlor may appear to be registered as 20 different businesses because 20 massage therapists with different names are registered to a single address. In reality, this should only count as one business since it has only one address. Therefore, duplicate records were identified and removed using PINS/addresses.

Another issue that raised is that some single businesses target more than one industry, resulting in multiple rows for the same business. In such cases, the business should be counted towards both industries.

**We have used Microsoft Excel to remove duplicates based on the constraints given.**

1. **Data Visualization:** Once the data was cleaned and preprocessed, Power BI was used to create interactive visualizations that provide insights into the industry mix/diversity on each commercial corridor. These visualizations included charts and graphs that show the distribution of industries across each corridor, as well as maps that highlight the location of businesses in each industry. Users could interact with these visualizations to explore the data in more detail and gain insights into trends and patterns.

The use of Power BI allowed us to effectively communicate the findings of the project to aid in strategic planning activities related to investments in and around these areas of the City of St. Petersburg.

# Results

Below table shows the industry mix/diversity across each of the eight commercial corridors analyzed in the project. Each row represents an industry, and each column represents a commercial corridor. The percentages in the cells represent the proportion of businesses in that industry located in each commercial corridor.

Graphical user interface, Excel

Description automatically generated with medium confidence

**Here are some key observations and insights regarding each corridor:**

1. **16th St Corridor:** The top industry category is **Retail Trade**, accounting for 29% of the total businesses. Other Services and Real Estate and Transportation and Warehousing are also significant contributors, accounting for 25% and 12.5%, respectively. The diversity is **lowest** as it’s restricted to 9 different industry categories.

Chart, pie chart

Description automatically generated

1. **18th Ave St Corridor:** The most prominent industry category on this corridor is **other services**, accounting for 25% of the total businesses. Retail trade and Real Estate and Rental and Leasing are also relatively high, at 22% and 16%, respectively. This corridor has a **moderate diversity** of industries, with businesses from 11 different industry categories.

Chart, pie chart

Description automatically generated

1. **22nd St S Corridor:** Most businesses on this corridor are in the **Real Estate** and Rental and Food services categories, accounting for 16% and 14%, respectively followed by the retail sector. This corridor has a relatively **high diversity** of industries, with businesses from 16 different categories.

Chart, pie chart

Description automatically generated

1. **34th St Corridor: Retail Trade** is the top industry category in this corridor, accounting for 25% of the total businesses. The other services sector is the next significant contributor, accounting for 34%, followed by food services. This corridor has a **high diversity** of industries, with businesses from 16 different categories.

Chart, pie chart

Description automatically generated

1. **49th St Corridor:**

The top industry category in this corridor is **Other Services**, accounting for 25% of the total businesses. The retail trade sector holds as the second highest contributor with 18%. Food Services and Real Estate and Rental and Leasing are also relatively high, at 10% and 8%, respectively. This corridor also has a **high diversity** of industries, with businesses from 15 different categories.

Chart, pie chart

Description automatically generated

1. **5th Ave Corridor:**

**Food Services** is the dominant industry category in this corridor, accounting for 32% of the total businesses followed by the other services sector. Construction and Real estate sector are also significant contributors, accounting for 10.53% each. This corridor has the **low** **diversity** of industries among all corridors, with businesses from only 10 different categories.

**Chart, pie chart

Description automatically generated**

1. **Central Ave Corridor:**

The most prominent industry category on this corridor is **Retail Trade**, accounting for 15% of the total businesses. Other, Professional and Health Care are also relatively high, at 14%, 13% and 12%, respectively. This corridor has the **highest** **diversity** of industries, with businesses from all the 19 different categories.

Chart, pie chart

Description automatically generated

1. **Dr MLK Jr St Corridor:**

**Retail Trade** is the highest contributor with 33%, followed other services, food sector and Real estate accounting for 16%, 14%% and 14% of the total businesses, respectively. This corridor has a **low** diversity of industries, with businesses from 9 different categories.

Chart, pie chart

Description automatically generated

# Conclusions

The analysis of industry mix/diversity on each commercial corridor indicates that each corridor has a unique mix of industries. The 16th St Corridor has the highest percentage of retail trade and other services, while the 18th Ave S Corridor has a high percentage of real estate and rental and leasing industries. The 22nd St S Corridor has a high percentage of unclassified industries, and the 34th St Corridor has a high percentage of food services and retail trade.

The 49th St Corridor has a diverse mix of industries, with high percentages of retail trade, other services, real estate and rental and leasing, food services, and health care and social assistance. The 5th Ave Corridor has a high percentage of food services, while the Central Ave Corridor has a high percentage of retail trade and professional, scientific, and technical services.

Finally, the Dr MLK Jr St Corridor has a high percentage of retail trade, food services, and health care and social assistance. The project sponsor should be aware of the unique industry mix of each corridor and how it might impact the development and growth of businesses in the area. Further analysis could be conducted on the specific types of businesses within each industry and how they contribute to the overall economy of the corridor.

One potential area of analysis is to explore the relationship between the industry mix/diversity and the socioeconomic characteristics of the neighborhoods in each commercial corridor. This could help identify any disparities or inequities in the distribution of industries across neighborhoods and inform policy decisions aimed at promoting economic development and reducing disparities. Additionally, it may be valuable to conduct a comparative analysis with other commercial corridors in the region to identify patterns and trends in industry mix/diversity.

# Link to Visualizations

Dataset Used: [Business Tax ID Data 2023.xlsx](Business%20Tax%20ID%20Data%202023.xlsx)

Access the Power BI report Here: [Power BI Report](https://app.powerbi.com/groups/me/reports/785eb156-6bc6-48f4-abf3-a6896bb884a7/ReportSection)