Introduction

The primary objective of this project is to analyze the data and to promote reinvestment in commercial corridors in the city of St. Petersburg. For this project, we analyzed new businesses that moved onto each corridor since 2015 and the number of new businesses moved onto any corridor each year. The analysis will help us concentrate on the corridors that have the least investments.

Methodology

The datasets given to us are **New Business datasets** from 2015 to 2022. These datasets contain information about new businesses like their location, which commercial corridors they belong to, postal code, Control Number, Business Name and description, and PIN of the address, for each year. Entry in these data sheets indicates a new business. We were given 8 Excel spreadsheets for each year from 2015 to 2022.

1. **Cleaning the Data**

These spreadsheets had to be filtered as they had a lot of formatting issues, null values, and inconsistency in the data. So, we consolidated all 8 sheets into a single sheet and filtered the data according to each corridor, one after the other. This gave us information on new businesses in a particular corridor. For example, we first filtered for 16th St Corridor and copied the results to a new sheet, and removed duplicate values based on Control Number. In the new sheet, we **introduced 2 new columns** namely **Corridor and Year**. The corridor column contains the corridor that the business belongs to, and the year column contains the year in which it was started. This addition was necessary to use these columns as attributes in Tableau. We noticed that the corridor names were not consistent throughout the given data. For example, the 16th St Corridor had different notations like 16th St Corridors and 16th St Corridors. We cleaned this

data and made it consistent for all the fields. We followed this method for all corridors and finally, we obtained a single filtered, consolidated spreadsheet. We formatted the Control Number column as a few entries had up to 4 decimal places, which is incorrect. We changed the format of this column to a number, with 0 decimal places. With this, our dataset was ready to use, and we named it **ConsolidatedDataforNewBusiness.xlsx**.

2. Using Tableau for our Visualizations

With the dataset that we created; we used Tableau to answer the given problem statements with Visualizations. Using bar charts was our best choice as it provides a precise and clear way of displaying our results.

Results

We used Tableau for the data analysis. We used bar graphs and a pie chart for the visualizations as the bar chart will provide a clear picture of the data in a precise manner, and the pie chart is easy to show the proportion of the data in each year.

The following is the list of research questions, their insights, and their visualizations.
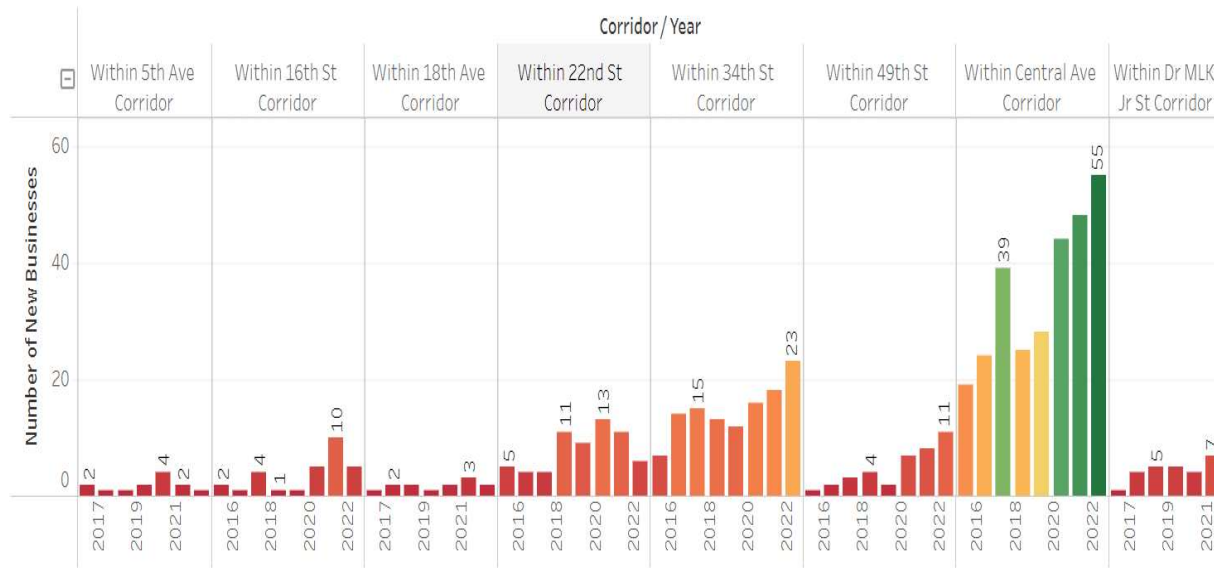
1. How many new businesses have moved onto each commercial corridor since 2015? Please provide an annual breakdown for each corridor.

For this question, we created a bar chart by taking the count of control numbers in rows and corridors and years in columns. We used a hierarchy for corridors and years to specify the values for each year in all corridors since 2015.

From the analysis, there is a significant count of the number of businesses added for the corridors 34th St Corridor and Central Ave Corridor compared to other corridors. The least number of businesses were added to the 5th Ave Corridor and 18th Ave Corridor. For a few corridors when there were 0 new businesses added, in a particular year, that year is not displayed in the bar chart. If a year is missing in the chart, it means that 0 new businesses were added in that year. For example, there is no value in the year 2016 for the corridor 5th Ave indicating no new business was added in that particular year. These observations reflect the corridors where there is a need for reinvestment.

Total Number of New Businesses Per Year in Each Corridor

We can view the data of corridor, year and count of new businesses added when we hover on the bar in the bar chart.

2. **How many new businesses, in total, moved onto ANY commercial corridor since 2015? Please provide an annual breakdown.**

For this question, we created a bar chart by taking the distinct number of new businesses in rows and years in columns. From the analysis, we can see there were a greater number of new businesses added in the years 2020, 2021, and 2022. The highest number of businesses (105) were added in 2021. We can handle the duplicates (new businesses that have more than one corridor) by considering the feature **distinct count** which will count only the unique values in the Control Number column.
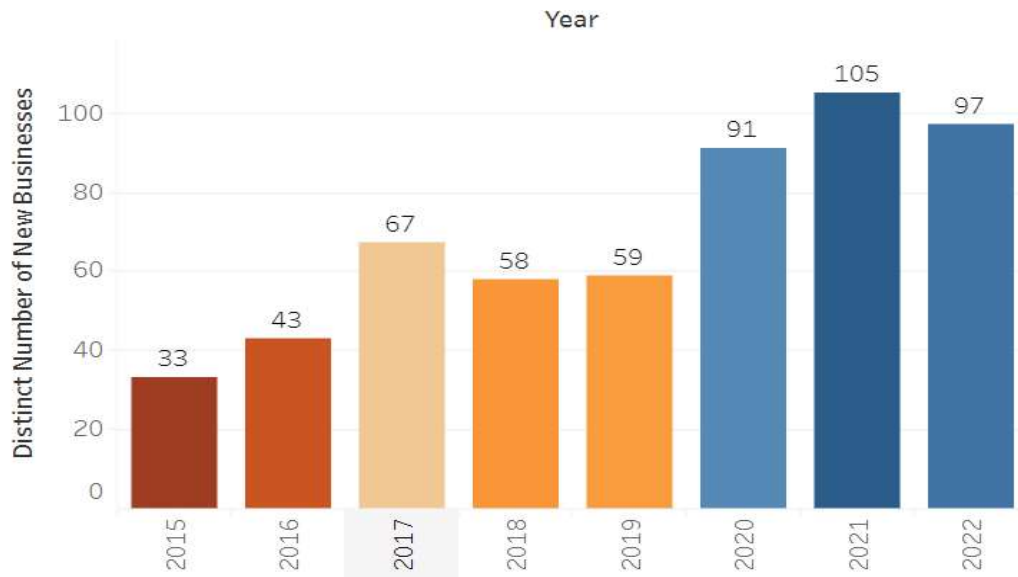
We can view detailed information like the number of new businesses and the year in which they were added, by hovering over a bar in the bar chart.

Conclusions

From the visualizations, we can conclude that the highest number of new businesses were opened on Central Ave. Corridor. This corridor has seen a gradual increase in new businesses being introduced, year by year, with the highest number of 55 new businesses recorded in the year 2022. 5th Ave Corridor and 18th Ave Corridors have seen the least number of businesses being opened on them. Based on these observations, we can conclude that Central Ave. Corridor has a high chance of new businesses being added to it. A few more measures can be taken with respect to the corridors that have the least number of new businesses, to make them more efficient and profitable. The reasons why this is happening can be investigated and reinvestment opportunities can be reconsidered.

We can also see that there has been a gradual increase in the total number of new businesses from 2015 to 2022. This indicates that there is great potential for more new businesses to emerge in the coming years.

Please use the link below to navigate to our Visualizations:

https://public.tableau.com/app/profile/rishika.nidiganti/viz/Group-7/Group-7#1

This link will take you to our Tableau Public account page which has our Dashboard. This link has 3 visualizations and 1 dashboard. The Dashboard is named Group 7 and it has all the visualizations that we created in it. By hovering the cursor on each value, detailed information can be obtained.

Appendix

For question 2, for total number of new businesses moved onto ANY commercial corridor since 2015, we can also view the results using a pie chart for displaying the percentage growth in the number of new businesses each year.

Proportion of New Businesses in Each Year
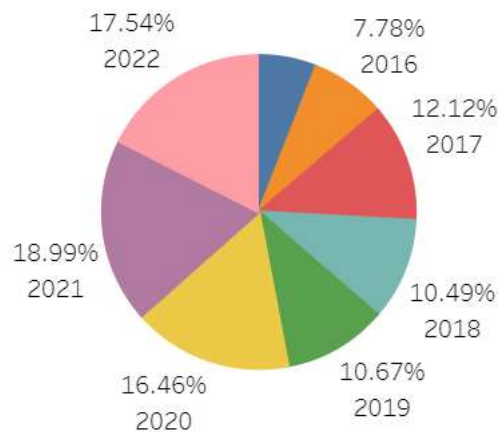
From the pie chart we can say that there is more contribution i.e., a higher number of new businesses were added for the years 2020, 2021, and 2022. The highest contribution (18.99%) was in the year 2021 compared to other years.