Data Warehousing Report

External Data Source Description:

1.Grocery stores in Seattle
https://data.kingcounty.gov/Health/grocery/8bgz-hyjq/about_data

This dataset provides a detailed account of grocery stores in Seattle,containing detailed information such as Name, Address, City, Zip Code, and Phone number.

It covers the period from 2006 to the present, offering an in-depth view of the grocery store landscape within the area. Designed to enhance our Seattle travel recommendation app, this dataset enables us to guide users towards nearby grocery shopping options based on their accommodation's Zip Code, enriching their travel experience.

Usage:

This integration will enable us to provide users with valuable insights into the availability of grocery stores near their chosen accommodations based on Zip Code. This feature is especially designed to enhance convenience for travelers looking to purchase groceries during their stay.

Furthermore, by analyzing the correlation between the density of grocery stores and food service establishments within various Zip Codes, we aim to uncover patterns that could inform travelers about the local food landscape, thereby improving their overall travel experience in Seattle.

2. Parks and green spaces in Seattle
https://data.seattle.gov/Parks-and-Recreation/Parks-and-green-spaces-in-Seattle-by-zipcodes/5uww-n58z

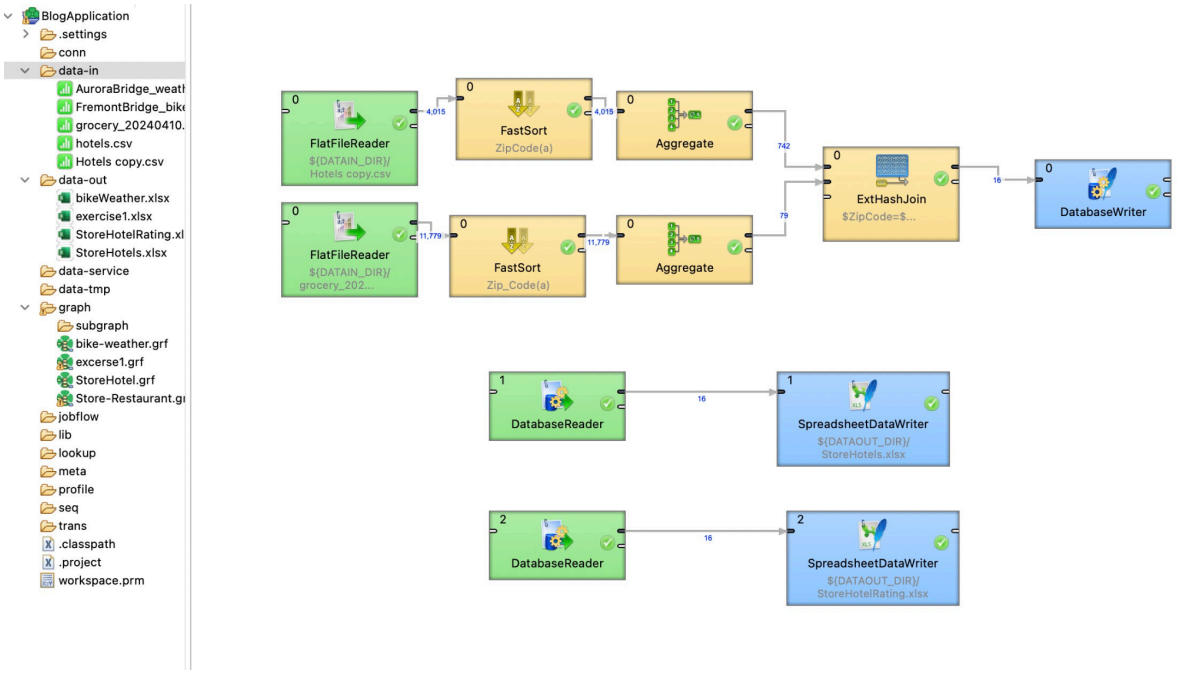
This dataset lists parks and green spaces in Seattle, cataloging each with unique identifiers (PMAID and LocID), names, and street addresses, along with ZIP codes for area reference. It also includes geographical coordinates (latitude as Y Coord, longitude as X Coord), and a combined location format for mapping purposes, making it a practical resource for geographical analysis and urban planning.

Usage:
Visitor-Friendly Analysis: The dataset could be used to determine which areas might be more visitor-friendly based on the presence of green spaces. It's commonly understood that areas with more parks are often better maintained, have more commercial activities, and potentially lower crime rates. These factors combined make such areas more suitable for visitors to Seattle. You might cross-reference this data with maintenance records, commercial activity levels, and crime statistics to get a full picture of each area's visitor-friendliness.

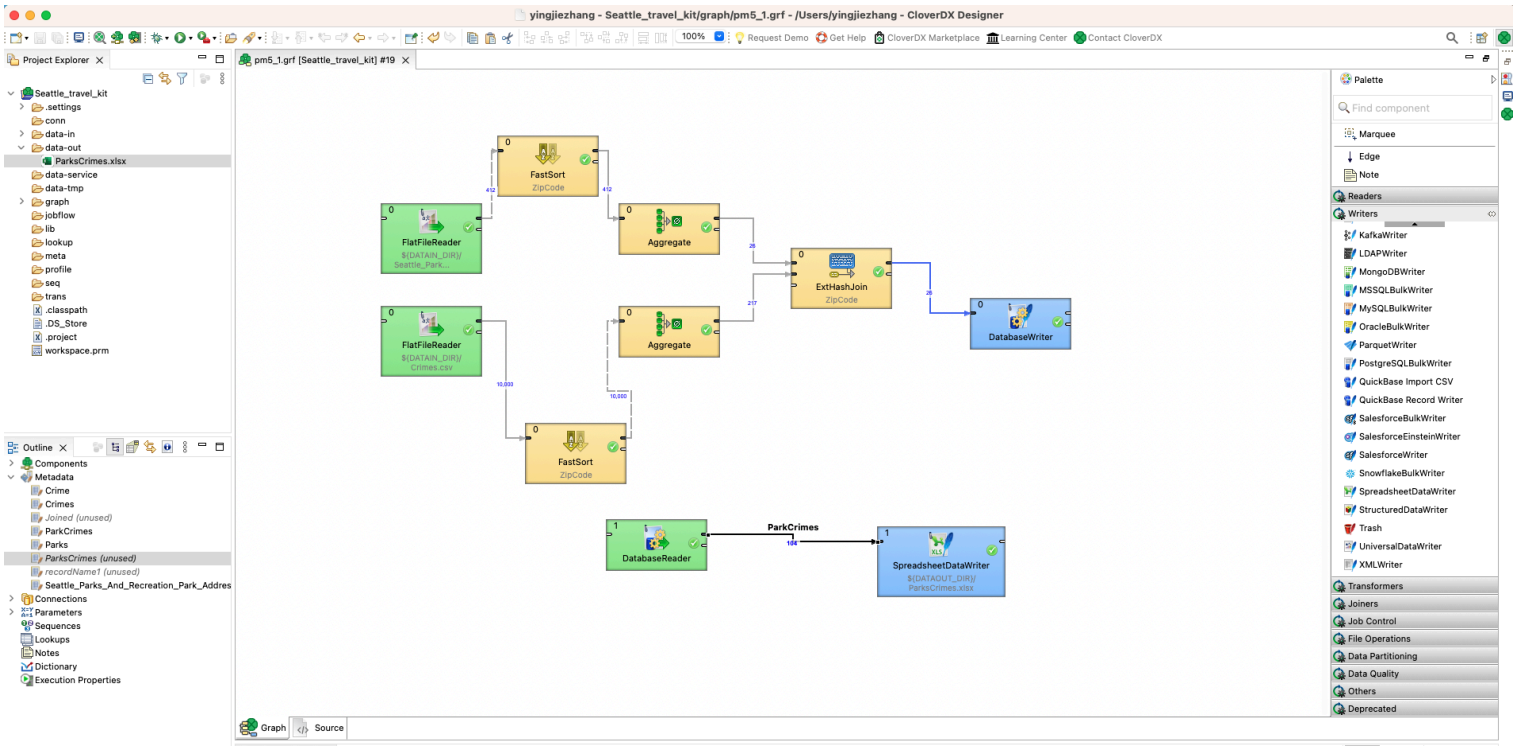
Zip Code Match: The ZIP codes can be matched with those of residents, tourists, or other areas to find the nearest parks or to analyze the distribution of green spaces across different neighborhoods in Seattle.

ETL descriptions

- each ETL: components used and what they do



FlatFileReader - Reads data from CSV files. There are two FlatFileReaders to read data from the hotels.csv and stores.csv.
FastSort - Sorts the data based on the 'ZipCode' field. This is done to prepare the data for the Aggregate component, which works more efficiently on sorted data.
Aggregate - Use this to count the storeCount, hotelCount and calculate the average ratings under each ZipCode.
ExtHashJoin - Joins data from aggregated hotels and stores by ZipCode.
DatabaseReader - Reads data from a database. There are two instances of this component, likely pulling data from the database.
SpreadsheetDataWriter - Writes data to an Excel file. There are two of these in the graph, indicating that the data is being written to two separate spreadsheet files.
DatabaseWriter - Writes data to a database. The output of the ExtHashJoin is being written to a database, to store the combined and aggregated results.



FlatFileReader - Reads data from CSV files. There are two FlatFileReaders to read data from the Seattle_Parks_And_Recreation.csv and Crimes.csv.
FastSort - Sorts the data based on the 'ZipCode' field. This is done to prepare the data for the Aggregate component, which works more efficiently on sorted data.
Aggregate - Use this to count the CrimeCount and ParkCount under each ZipCode.
ExtHashJoin - Joins data from aggregated crimes and parks by ZipCode.
DatabaseReader - Reads data from a database, pulling data from the database.
SpreadsheetDataWriter - Writes data to an Excel spreadsheet file.
DatabaseWriter - Writes data to a database. The output of the ExtHashJoin is being written to a database, to store the combined and aggregated results.

Chart descriptions/conclusions

Chart 1: Grocery stores & restaurants

Hypothesis for combining the data:
We hypothesize that in Seattle, the number of grocery stores and restaurants within the same Zip Codes are positively correlated, indicating that busier areas tend to have more of both.

Results of combining the data:

ZipCode	StoreCount	RestaurantCount
98002	407	12
98003	296	44
98004	195	61
98005	113	5
98006	162	1
98007	225	17
98008	46	1
98011	110	9
98023	202	10
98027	211	4
98028	101	5
98030	217	11
98031	134	10
98032	245	36
98033	124	45
98034	289	18
98038	104	3
98040	97	4
98042	175	1
98052	462	32
98056	152	3
98057	262	9
98072	112	23
98077	38	1
98101	156	28
98102	80	11
98103	241	20
98104	131	12
98105	169	6
98106	66	1
98107	255	5
98108	149	2

Chart's significance for application and the Action could take:
The data presented in the chart validates the hypothesis that in Seattle, the number of grocery stores and restaurants within the same zip codes is positively correlated. This suggests that areas with a higher density of such amenities are busier and potentially more attractive for both residents and visitors. For a travel application, this correlation is significant as it indicates bustling hubs of activity that are likely to offer a variety of dining and shopping options—key points of interest for travelers.

Action could take: Enhance your travel app by featuring a "Local Hotspots" section that showcases areas with high numbers of grocery stores and restaurants. Promote these vibrant neighborhoods as must-visit destinations, providing users with a curated list of popular spots and exclusive deals.

Chart 2: Grocery stores count & Hotels count

Hypothesis for combining the data:
Within the same Zip Codes in Seattle, we assume that the quantity of grocery stores positively correlates with the number of hotels. The reasoning is that for areas with more hotels, there will be higher tourist foot traffic, thus companies are more motivated to grocery stores in the same area. Thus Zip Code with more hotels tend to have a greater number of grocery stores.

Results of combining the data:

The data visualization appears to show a weak correlation between the number of grocery stores and the number of hotels in the same zip codes in Seattle, especially in zipcodes around 98100 - 98105 since that's indeed the zipCode for downtown Seattle. The data supports our hypothesis.

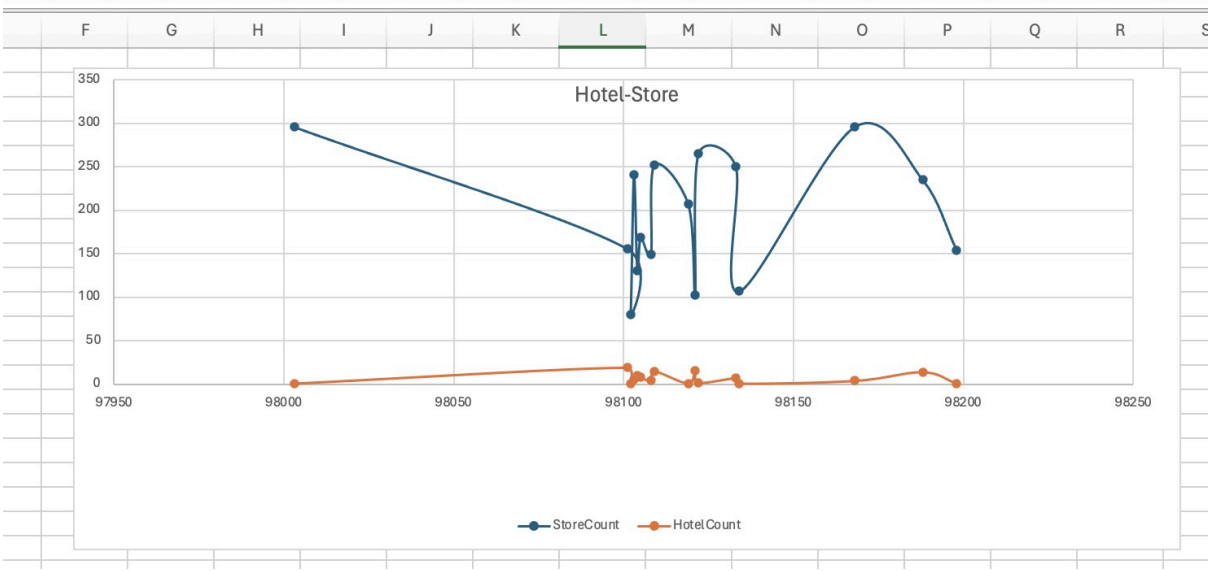


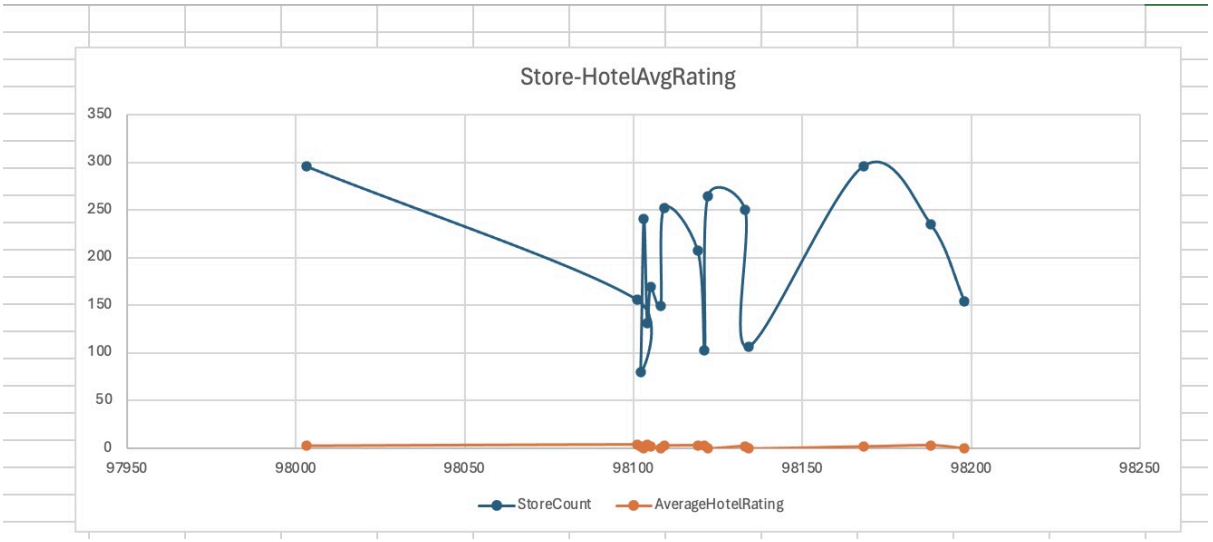
chart: relation of hotelCount and storeCount

Chart's significance for application and the Action could take:
For our SeattleTravelKit application, this correlation can be used to enhance user experience and service offerings. Recognizing that tourists in hotel-rich areas may seek convenient access to grocery stores for daily necessities or local specialties, the application can:
(1) Consider adding a feature to list the grocery stores in areas with high hotel density, as these listings may receive more traffic and use by tourists. For example, we could map out grocery stores in proximity to hotels, allowing tourists to easily locate nearby shopping options.

(2) Besides listing, allow users to comment on shopping, leveraging the interplay between accommodation and retail to enrich the travel experience in Seattle.

Chart 3: Hotel Average Rating & Hotels count
Hypothesis for combining the data:
In Seattle, a higher average hotel rating within a Zip Code may attract more hotel establishments, suggesting a correlation between hotel quality (hotel average ratings) and count.

Results - relation of storeCount and AverageHotelRating

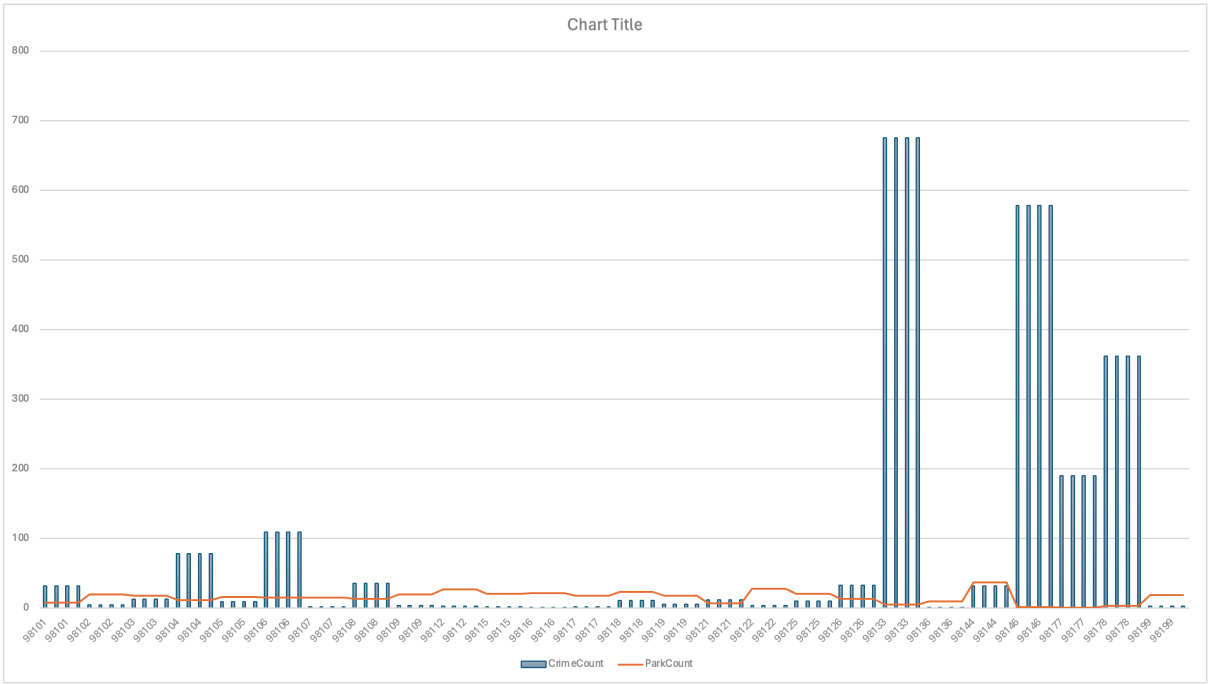


Chart's significance for application and the Action could take:
The data provides insights into how the quality of hotels (average rating) may influence their quantity in specific areas, which can be valuable for travelers looking for the best experience. Higher-rated hotel areas might be more desirable to travelers, indicating better services, amenities, or safer neighborhoods. The application can allocate more resources to list and detail hotels in Zip Codes with higher hotel counts and ratings, assuming these are prime locations for travelers.

Enhance the travel app to highlight areas with a high count of highly rated hotels, possibly suggesting a better overall visitor experience. Offer personalized recommendations for users by matching their preferences with areas that have a high density of quality hotels. Partner with high-rated hotels in areas with a greater hotel density to offer exclusive deals, capitalizing on the draw of these locations.

Chart 4: Green Space & Crime
Hypothesis for combining the data:
Neighborhoods with more and well-maintained green spaces may experience lower crime rates, as these areas foster community engagement and vigilance that can deter criminal activity.

Results of combining the data:



zz

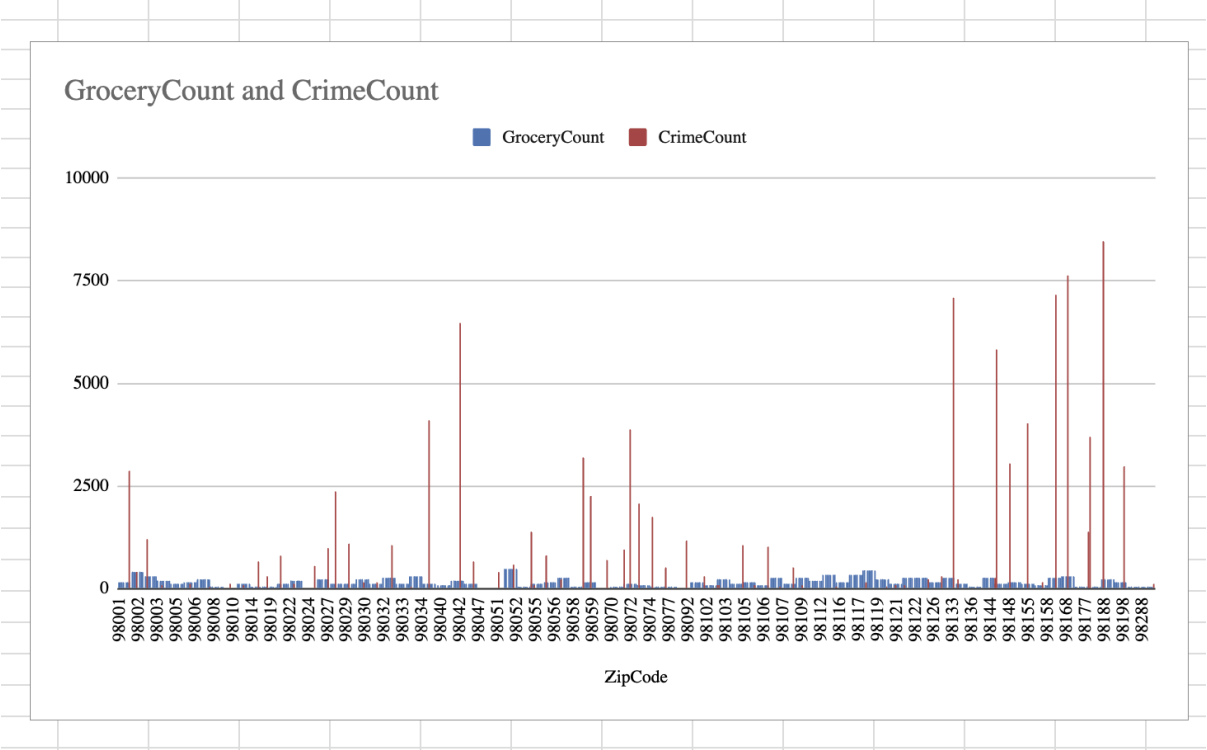
Chart's significance for application and the Action could take:
The data from Chart 4: Green Space & Crime highlights the positive correlation between well-maintained green spaces and lower crime rates in neighborhoods. This insight can enhance the appeal of destinations featured in your travel app, reassuring users of their safety and encouraging them to visit and explore these areas. The result validates our hypothesis.

Action could take:

Add a "Green Safe Spots" Feature: Based on the findings from Chart4 that link well-maintained green spaces with lower crime rates, we can integrate a new feature into the travel app called "Green Safe Spots." This feature will highlight neighborhoods with abundant green spaces that are known for their low crime rates. Users can explore these areas with confidence, knowing they are in a safer environment. Offer detailed guides, safety ratings, and real-time user feedback on these areas. Enhance user engagement by including interactive maps that users can customize to include these green spaces in their travel itineraries, promoting not only safety but also the enjoyment of urban nature.

Chart 5: Grocery stores & Crime
Hypothesis for combining the data:
The number of grocery stores within an urban area may be inversely related to the crime rate, as a higher density of grocery stores could indicate a thriving local economy and a more active community presence, both of which are factors that can contribute to deterring crime.

Results of combining the data:
The data visualization appears to show a weak correlation between the number of grocery stores (GroceryCount) and the number of reported crimes. From zipcode 98102 to 98133, we observe a trend that a higher number of grocery stores is associated with a low crime count, which contradicts our hypothesis. This indicates that a more active community presence does not necessarily bring a higher crime rate.



Chart's significance for application and the Action could take:
Add a feature "Safe convenient area". Based on the findings from Chart5 that a more active community presence does not necessarily bring a higher crime rate, we can integrate a new feature into the travel app that highlights/recommends areas with low crime rates and a high density of grocery stores. This provides users with suggestions on areas that are potentially both safer and more convenient for visitors.