**Abstract:**

Speedway Junk Removal is a company that I used to work for during my undergrad at the University of Arizona. We would go to houses in the greater Tucson area with a dump truck and remove items accordingly. The project uses the Speedway Junk Removal’s customer database that tracks job and customer data and compares it to the financial and geospatial relationship of Zillow’s Home Value Index the average housing price per zip code. Much of my theory when working as an operations manager at Speedway Junk Removal for nearly two years was nicer houses lead to higher quality job opportunities where homeowners have a greater disposable income for things like junk removal. According to the city of Tucson’s government website, “Each day, approximately 2,300 tons of solid waste is brought to the landfill for disposal. Los Reales is open to private, commercial haulers, and residential self-haulers" (Los Reales). Speedway Junk Removal falls under the commercial side of this static and shows that there is plenty of financial gain in Tucson’s waste management services. Building models that allow business owners to correlate different influencing aspects that drive business such as home value to revenue reinforces how marketing decisions are made. This could lead to further analysis and data outcomes of Speedway Junk Removal’s customer database to show where the company could focus their door hanger marketing strategy as well as areas of Tucson that are unfavorable.

**Table of Contents:**

*Abstract*

*Problem Statement*

*Connection to the Reading*

*Possible Data Outcomes*

*Data Preparation and Analysis*

*- Data Sets*

*- Data Collection*

*- Cleaning Data*

*Results*

* *Bar Graph 1: Average Job Revenue by Zip Code*
* *Bar Graph 2: Average Job Revenue by Zip Code (10 or more Jobs)*
* *Bar Graph 3 Average Home Value Index*
* *Scatter plot 1: Linear Relationship between Home Value and Job Revenue*
* *Map 1: Raster Map of Zillows Home Value Index by Zip Code*
* *Map 2: Raster Map of Speedway Junk Removal’s Revenue Distribution by Zip Code*

*Discussion of Future Applications:*

*References:*

**Problem statement:**

The goal of this project is to determine the geospatial relationship between Speedway Junk Removal’s job revenue and Zillow’s Home Value Index (ZHVI) per zip code in Pima County. Societies actions often correlate with financial purchasing power and can lead to incites on customer decision making. By combining these three datasets Speedway Junk Removal will have a better understanding of the forces that drive revenue in certain zip codes in the greater Tucson area. Particularly when taking into account marketing strategies like door-to-door sales and door hanger flyers.

**Connections to the literature:**

Data is everywhere, even in junk removal; it is important to be able to perform analysis on the data we collect as it can lead to key incites for business decision making. In “Be Data Literate: The World of Data” by Jordan Morrow explains the importance of data literacy and stresses the usefulness of data informed decision making rather than strictly data driven decisions; by learning these data skills it closes the skill gap between the work force and these data incites.

**Outline of Possible Outcomes:**

To create a usable model to track revenue by zip code in Tucson’s Pima County and how that relates to the average home value in that same zip code. The models' elements will consist of a scatter plot of each zip codes revenue in relation to Zillow’s Home Value Index. The Pandas data frame allows for reusable code when further jobs are recorded in the future. Data will be written to a csv and downloaded in QGIS to build a model that maps the data for both the average home value, and Speedway Junk Removals revenue distribution.

**Data preparation and analysis:**

Data Sets:

*Data Set 1****:*** Zillow zip code average Housing prices across the country since 2000.

*Data Set 2*:Speedway Junk Removal HouseCall Pro Customer Database

*Data Set 3:* Pima County Shapefile with zip code boundaries and labels

Data collection**:**

Housing data was taken from Zillow’s research website and selected to aggregate by all US zip codes. This provides an average home value index based on the houses in that zip code. Since 2000 Zillow has calculated a monthly average home value index which is “a measure of the typical home value and market changes across a given region and housing type. It reflects the typical value for homes in the 35th to 65th percentile range” (Housing Data). Zillow has filtered many of the outliers for low- and high-end homes in that same region by focusing on the median percentile valued homes. More information on exact calculations can be found on [Zillow’s research user guide.](https://www.zillow.com/research/zhvi-user-guide/)

Speedway Junk Removal’s data is collected on [Housecall Pro,](https://www.housecallpro.com/) a service application that automates job communication among customers and employees. When a new job is either placed online or over the phone a customer's address is documented as well as name, phone number, job description and total job revenue after payment is taken. Housecall Pro allows small businesses to create and track customer information in a database by logging important job data for future reference. Data is collected by owners and managers who work at Speedway Junk Removal. Addresses are verified in google maps and revenue data is taken directly from the payment history on HouseCall Pro.

Cleaning Data:

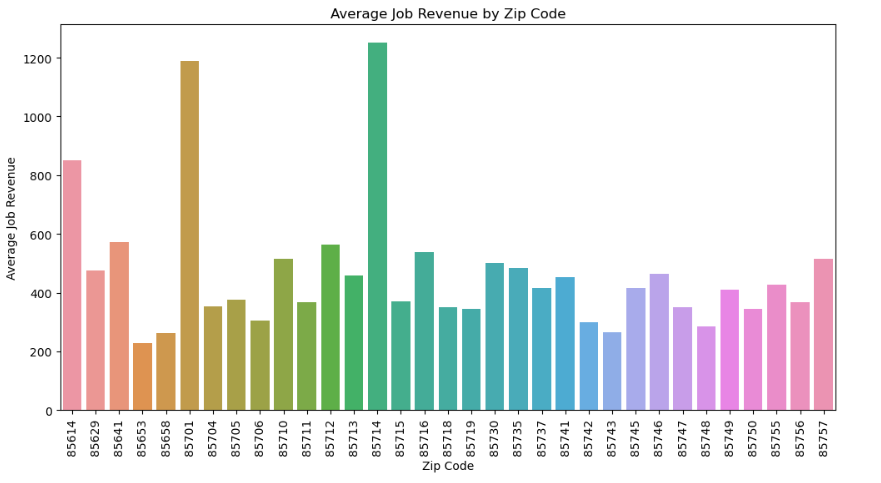
Each data set was downloaded and converted to a Panda's data frame to aggregate and clean data. For the purposes of this research project, we chose to focus on the zip code column in all three data sets, which will act as the primary key of cross reference for each aggregation. For geocoding purposes and future outcomes, a column was created in the Speedway Junk Removal (SPJR) data frame for concatenating all portions of the customer address, allowing for precise location when mapping. The zip code column was checked for total unique values to find the maximum number of different zip codes that jobs have recorded revenue in (27). Working with the 27 different zip codes a panda’s group\_by aggregation function was preformed to find the average job revenue by zip code instead of per job. The agg function also gives the total count of jobs recorded in each zip code and a standard deviation measure for average revenue. The count allows for proper filtration of the data only using the zip codes where the count is greater than 10. This will allow for a higher degree of accuracy in our average revenue calculation, eliminating some of the outliers in the data set where jobs are skewed for an unlikely job revenue. Jobs with zero revenue were also dropped to avoid underestimation.

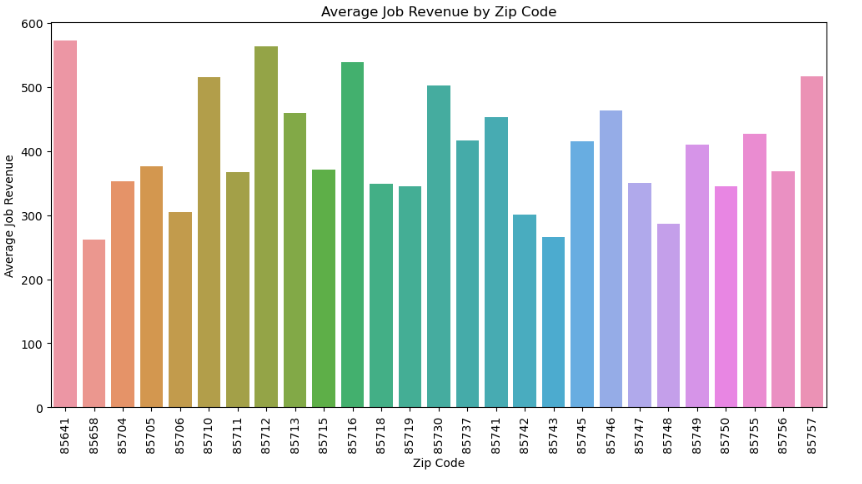
Zillow housing data provides housing data for all US zip codes. The data was filtered down by finding all zip codes in Pima County by searching for ‘Pima County’ in ‘Metro’ column of the Zillow data frame. This includes the city of Tucson and the surrounding areas. The Zillow Home Value Index is recorded on the last day of each month. This means we must average by row or “the axis =1” in pandas to determine the average Home Value Index for a zip code in our selected timeframe. Since Speedway Junk Removal opened in 2020, this provides are starting point for the timeframe selected in this project. The average was calculated for the four-year time span and the result was placed in a new column called “Average Home Value”. Finally, the Zillow Home Value Index data frame was merged with Speedway Junk removals customer database by the primary reference key (zip code). Some of the Zip codes in the Zillow data frame were dropped because they did not have any recorded job visits.

Once the final data frame cleaning edits were made, the new data frame was written to a csv and opened in QGIS. Where further raster map visualizations were made after a join between the final data frame in pandas and a geospatial shapefile taken from ArcGIS categorized by zip code. This join was performed in QGIS and allowed for raster maps of both home value and revenue.

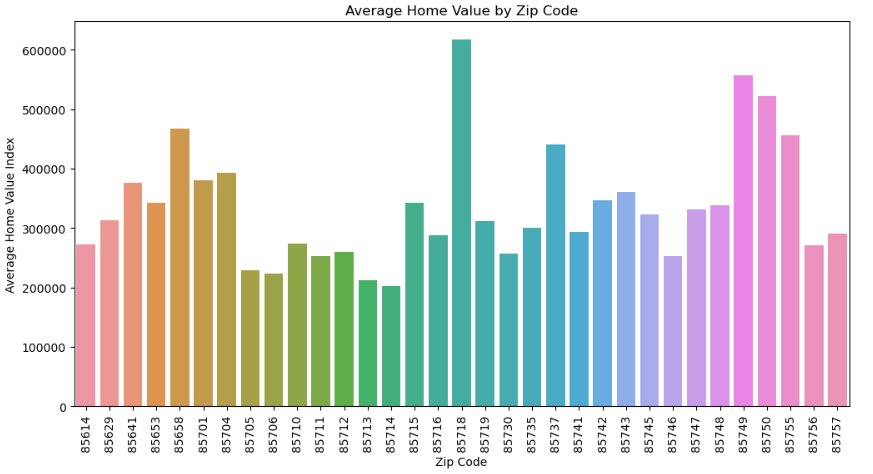
**Results:**

The results show that home value and total revenue collected in that zip code have a slightly negative relationship and no correlation. Meaning that home value doesn't directly affect the amount of revenue a job can place based on its geospatial relationship. Even when filtering out the outlier zip codes the distribution for the scatter plot seemed to have no correlation. After calculating Pearsons R correlation coefficient for the scatter plot (-0.29) this shows that there is likely no linear relationship among the two variables and negative one at that. Home value is usually correlates to purchasing power, however with such a small sample size of jobs with recorded revenues (just over 1200) this still might skew the zip code regions.

Graph 1 The average job revenue for all zip codes visited by Speedway Junk Removal with no limit on the count for zip codes visited. (Every Job with Revenue)



Graph:2 Removes Outlier zip codes by eliminating zip codes where the count (or total number of job visits) is less than 10. This provides a greater strength in the averages produced for each zip code.



Graph 3: Shows Zillow’s Average Home Value Index by zip code.

Scatter Plot 1: Shows the negative linear relationship between the Average Home Value Index and job revenue in the same zip codes from 2020 to 2024. The slope is negative (-0.000248) showing a negative relationship and fairly no correlation determined by our correlation coefficient (R value of –0.29). The probability of getting this outcome determined by our P-value (0.139) determines that outcomes are statically likely and that there is little to no correlation.

Map 1: The geospatial distribution of home value in Pima County. With the maximum average home value being over 600,000 and the minimum above 200,000. The raster map shows the change in home value by decreasing color.

Map 2: The geospatial distribution of job revenue in Pima County for over 1000 recorded jobs. Darker Areas mean higher average job revenue is collected in that zip code for SPJR.

Further results:

After mapping the data, we can see that the higher valued homes are towards the northeast, these zip codes are towards the Tucson foothills which is one of the higher end parts of Tucson. However, when examining job revenue much of the higher revenue zip codes are in the center of Tucson where Speedway Junk Removals facility resides as well as to the South where newer housing developments are taking place. The distribution of revenue for a single job mostly depends on the amount of stuff being removed, since volume of the truck was one characteristic for determining price of a job.

**Discussion:**

By using the data, we collect we can build models for businesses to help make informed decisions, to complete this model a web scrapping portion for both data sets could be utilized and implemented to update model when a job is added to the customer database same thing could be said for Zillow's Home Value Index. Other factors such as total volume and job description play a more vital importance in determining total revenue for a job. If the customer database recorded the date of each job, we could perform an analysis for revenue over that period to show the growth or fall in business every year.

**References:**

*Housing Data*. Zillow. (2024, September 4). <https://www.zillow.com/research/data/>

*Best junk removal in Tucson AZ: Junk Hauling & Disposal*. Speedway Junk Removal. (2023, July 31). <https://speedwayjunkremoval.com/>

*Los Reales Sustainability Campus*. City of Tucson. (n.d.). <https://www.tucsonaz.gov/Departments/Environmental-and-General-Services/Los-Reales-Sustainability-Campus#:~:text=Each%20day%2C%20approximately%202%2C300%20tons,%2C%20and%20residential%20self%2Dhaulers>.

Arcgis.com. (n.d.). <https://www.arcgis.com/home/item.html?id=7401dd1ccfea4ff5901092324f8b44ed>