An Analysis of Real Estate Transactions Using Zip Codes and Office Locations

Eden M. Anderson

Northwest Missouri State University, Maryville, MO 64468, USA
S544978@nwmissouri.edu and eden.anderson@betteromaha.com

Abstract. Due to the fact that the real estate industry has a strong impact on both individuals (micro) and the local economy (macro), a real estate brokerage needs to be strategic in order to achieve success within their market. This project aims to determine if the location of a real estate brokerage office has an impact on the location of a brokerage's real estate transactions. Data from a real estate brokerage in the Omaha, NE, metropolitan area was collected for 2022. This data was used to compare the zip codes for each closed transaction to the office location of the Realtor representing the transaction. The data set of 3140 records was trained and tested on three Classification models (K Number Neighbors, K Means Cluster, and Random Forest). While the K Number Neighbors and Random Forest models were very similar, maximum accuracy was found using the Random Forest model. The statistical evidence shows that, while the models do predict the zip codes with about two-thirds accuracy, they do not have the reliability to be used for business decisions based on predictions. The information, however, could be considered useful to a brokerage for marketing and other lower-cost investment purposes; individual Realtors within a brokerage could also use these results to further develop their business plans.

Keywords: real estate · data analysis · zip codes · Realtor · office location

1 Introduction

Real estate is an industry that has an impact on nearly every person, as most people are concerned with making sure they have a place to live. Whether housing consists of a single family unit, a duplex, or even an apartment, people are invested in creating the best living situation. [9] Economies are also impacted by the housing market, making people even more invested in what is happening in the current real estate market. [16] As such, the role of a Realtor is integral in securing the best outcome for those seeking to move from one housing unit to a different one. [12]

This project will analyze the potential relationship between real estate transactions and a Realtor's physical office location. This project examines an important question for the owner(s) of a brokerage with several physical office locations: Does the physical office location of a real estate agent impact the location of their real estate transactions? Real estate transactions from a local brokerage will be used to compare the zip codes of a Realtor's transactions to that of their physical office location.

By analyzing this data, a broker is presented with information that can help them determine several key business factors: If there is a direct correlation, are more office locations needed? If there is a correlation, will real estate agents need to move physical office locations to better serve their businesses? If there is a correlation, can the brokerage provide more targeted support to an office branch - including, but not limited to, marketing and neighborhood involvement? If there is not a correlation, should a broker close an office branch?

This project will follow the steps of the Data Science Lifecycle as seen in Figure 1. [1] The first step of business understanding occurred when a project was identified that can help owners of a local brokerage to make business decisions regarding office locations. From there, specific data was identified for use; the data will come from the local brokerage and will contain information about real estate transactions that occurred in 2022. Next, the data will be reviewed for missing data, duplicates, and split data. This will be corrected, and the data will then be analyzed for possible correlations. The data will also be split into training and test sets at this time. Next, a Machine Learning algorithm will be used to construct and train a model using the training data set. The model will be adjusted and retrained as appropriate. The test data set will then be run through the model to evaluate and make predictions. Finally, a conclusion will be made, and the results will be reported using data visualization.

1.1 Goals of This Project

The goal of this project is to identify a relationship between real estate transactions and the physical office location of the Realtor involved in the transaction. By identifying this correlation, owners of a brokerage will be able to make business decisions regarding branch locations, number of branches, and Realtor placement throughout those locations. The results could possibly impact recruiting and marketing, as well as investment in office infrastructure.

1.2 Limitations

This project is potentially limited by the data being used, as it is from a local brokerage in Omaha, NE: The model built may not be transferable to another brokerage. In addition, this project is limited by the low-inventory and high-demand market that existed in Omaha, NE, in 2022: The shifted market could have affected where people were looking to purchase homes, taking them farther from a desired location in order to find a desired price. [7] Finally, this data is limited by the Omaha metropolitan area, and may not include nearby locations (i.e., Council Bluffs, IA, and Lincoln, NE). [15]

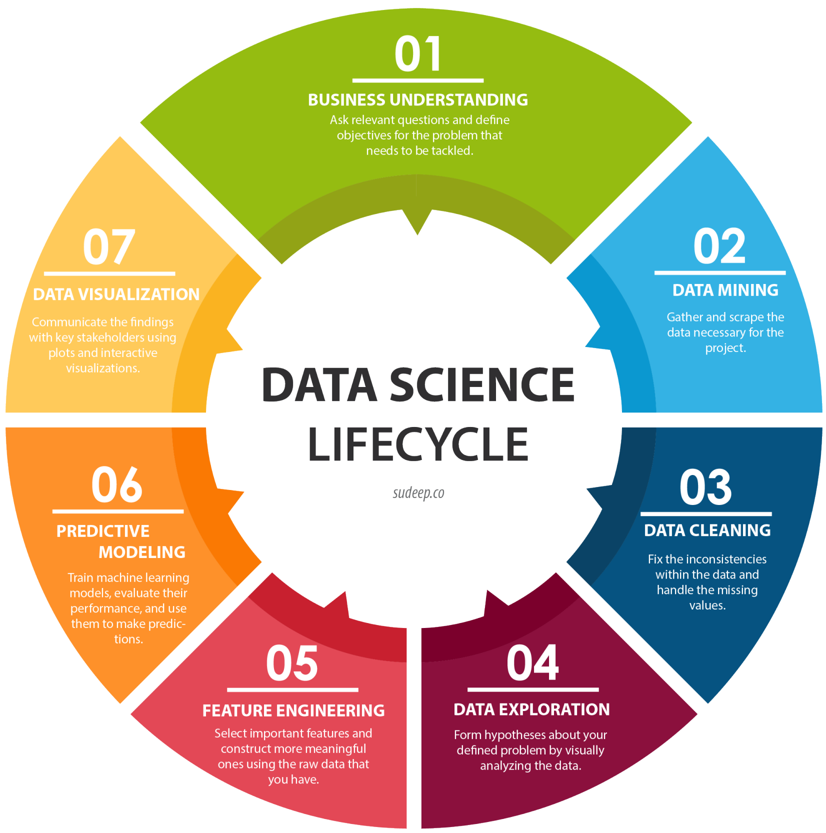

Fig. 1: Data Science Lifecycle by Sundeep Agarwal. This image is useful for understanding each phase of the data science life cycle. [1]

2 Related Work

There are three major topics discussed in this project: office (branch) location, Realtor choice in brokerage, and consumer choice in Realtor. The decision of where to build an office can impact any industry, as seen in the article by Görener et al. In the article on determining the optimal location when the location for a new bank branch location was being sought, they demonstrated the need for a process to measure three main criteria to address both banking and consumer requirements. In addressing this topic, the authors developed a technique that optimized then ranked potential locations. [10] The report by Görener et al. supports the hypothesis put forth in this report that an office location can have a direct impact on the surrounding area.

This report also considers the topic of how and why a Realtor might choose a specific brokerage. In a paper that discussed the increasing competition between real estate brokerages, Barwick and Wong introduced multiple factors that affect a Realtor's decision to join a specific brokerage. Barwick and Wong discussed technological innovations that had recently been introduced in the real estate industry; they also addressed competition and market share and how that impacted local brokerages. The paper mainly focused, however, on commission structures and the impact they have in a Realtor's selection. [8] Having real estate agents that generate success in both sales volume and units sold first impacts whether a brokerage might even need a new office location. That success would then help to determine where a new branch should be located to have maximum impact on both the Realtors and the consumers.

The third topic that this report addresses is that consumers have a choice when selecting a Realtor. In their report, Newell and Plank addressed two different kinds of relationship that a Realtor must have with a client: interpersonal and consultative. Each relationship would fulfill a consumer need that, when addressed, would create a long-lasting relationships. [13] Building long-term relationships can generate repeat business and also lead to new business through referrals. These relationships directly impact a Realtor's success.

3 Real Estate Transaction Data

The collection of this data was broken down into three steps: The first step consisted of identifying the data source and collecting the relevant data. The second step was preparing the data to be analyzed. The third and final step involved the actual data analysis through exploratory data analysis and predictive analysis.

3.1 Data Source and Collection

This project was intended to provide usable information for the brokers at a local real estate brokerage in Omaha, NE. Permission was given to use actual data from the company's 2022 production. The data was pulled from a secure, franchised website, so a direct link to the data source was not accessible to the public. The data can be found at this GitHub link [2].

The data set was composed of structured data derived from a database of closed real estate transactions from a local brokerage in Omaha, NE, that occurred January 1, 2022, through December 31, 2022. The data was originally downloaded as an Excel file. There were 3545 records in the data set with 12 fields, and the data set's file size was 321 KB. The data fields included Numeric Integer (Company, MLS Number, Zip/Postal Code, Sales ID), Date/Time (Actual Close Date), and Character String/Text (Transaction ID, Property Address, City, State/Province, Sales Volume, Local Currency). This data was very good: The data records were complete for the fields that were necessary for this project (Company, Postal Code, and Sales ID), and the data was formatted the same for each record.

3.2 Data Preparation

Because the data was imported from a verified source as an Excel file, the data was confidently known to be a complete set of real estate transactions for a local brokerage in 2022. Formatting of the OFFICE IDs then needed to occur to remove unnecessary descriptive text. The data also included personal identifying information that included Realtor names and commission amounts. These were removed to protect private information, as discussed with the brokers prior to their approval of using this information. Next, there were columns removed from the database that were unnecessary for this analysis (i.e., Listing Date, Estimated Close, etc.) In order to make the data more usable, the Excel file was then converted to a CSV file format to prepare for future analysis. The cleaned data can be found at this GitHub link [5].

At this point, the CSV file was evaluated to determine the data quality (i.e., validity, consistency, accuracy, precision, uniformity, and completeness). This included making sure that city names were spelled correctly and that all zip codes were five digit, which involved removing the extra four digits on 37 records. There was one record that had missing city, state, and zip code information, and this was looked up in the MLS (Multiple Listing Service) to add the correct, missing information. In this instance, using a technique to replace missing data with a median or mean of the attribute was unnecessary because accurate information for one record was easy to verify. (If the sales price had been a necessary attribute, using the median or mean values would have been prudent.) The CSV file was also used to remove transactions that occurred outside of the Omaha metropolitan area [15]. This reduced the number of files to 3140 records. A secondary CSV file was created to differentiate between the initial sourced data and the cleansed data.

Upon further review, the cleansed CSV file contained unnecessary attributes: The Company, Trans ID, MLS Number, Sales Volume, and Currency columns were removed because they did not impact the data to be analyzed; there was also repetitive and missing data within these columns. The headings of the remaining attributes were then formatted to make referencing them easier in Python:

Underscores were added between words. These final cleansing steps resulted in 3140 records and 7 attributes.

Preparing the data readied the major data attributes that would be used for the project. The following attributes remained in the data set:

- OFFICE - The Office attribute identified the home office of the Realtor.
- PROPERTY ADDRESS - The Property Address attribute identified the precise house number and street of a transaction.
- CITY - The City attribute identified the city of the transaction and was used to verify that the transaction occurred within the Omaha, NE, metropolitan area.
- STATE - The State attribute was associated with the Property Address and City attributes and verified that the address was in the state of Nebraska.
- ZIP/POSTAL CODE - The Zip Code attribute was associated with the Property Address attribute and was important in the analysis to determine if there was a correlation with the office location.
- ACTUAL CLOSE DATE - The Actual Close Date attribute verified that the transaction occurred during the correct time frame.
- SALES ID - The Sales ID attribute identified the main Realtor involved in the transaction and was used to determine if there was a correlation between the locations of the transactions and the office location.

Of these attributes, the independent variable was identified as OFFICE, which was used to evaluate the locations of the transactions. The dependent variables were the ZIP/POSTAL CODES of the transactions, which was compared to OFFICE, and the SALES ID, which was used to determine if the transactions occurred near an office location.

The CSV file was then imported into JupyterLab to further verify the quality of the cleansed data by using Python modules. It was confirmed that there were no missing or null values in the main attributes. The data types of the OFFICE, ZIP/POSTAL Codes, and the SALES ID attributes were confirmed as numbers. (dtype: int64)

The OFFICE attribute was evaluated using a boxplot visualization to show the distribution of the transactions. The boxplot visualization was used to verify that the OFFICE ID should be the dependent variable. (See Figure 2)

The SALES ID and ZIP/POSTAL CODE attributes were also evaluated to review the distribution properties. In these instances, however, a histogram better displayed the results. (See Figure 3 and Figure 4)

The SALES ID histogram showed that there were two separate groupings of internal Realtor IDs. These groupings were known to be the result of ID numbers created by two different software packages within the franchise, and the difference in numbers had no meaning or impact on the results. The ZIP/POSTAL CODE histogram showed a tight grouping of 5-digit numbers, which represented the zip codes of the Omaha, NE, metropolitan area.

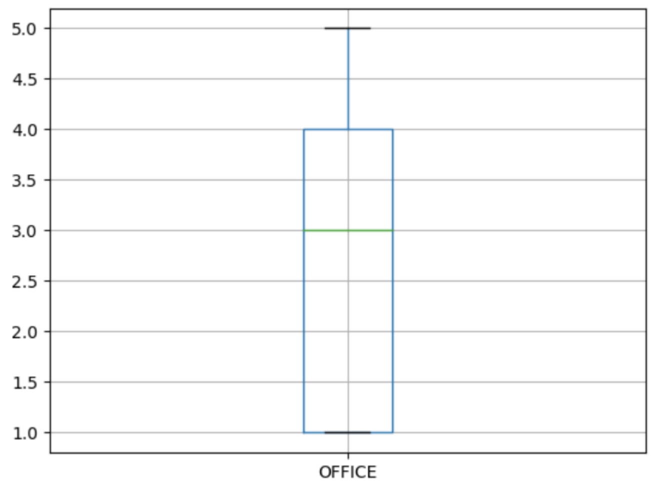

Fig. 2: OFFICE ID Boxplot. This image showed the distribution of the OFFICE ID, which helped to identify it as the independent variable.

The data cleansing process created a data set that was ready to be used for analysis. The independent variable (OFFICE) and the dependent variables (SALES ID and ZIP/POSTAL CODES) were identified and verified as quality data. These would be used to analyze the correlation between office locations and transactions of the Realtors that were based out of those offices.

An additional Excel file was downloaded and converted to CSV file format in order to provide information that matches SALES ID numbers to brokerage office locations (OFFICE). This file was formatted in a similar way: Personal identifying information was removed from the file prior to posting on GitHub. In addition, the OFFICE numbers were formatted to remove unnecessary descriptive text. This file was not used, but it was included for the extra data that might have been needed.

3.3 Data Analysis

Performing data analysis occured in two steps: exploratory data analysis and predictive data analysis. Exploratory data analysis (EDA) is the process of investigating the data set to summarize and analyze the attributes in order to gain a better understanding of what the data set has to offer. EDA often involves visualization to demonstrate any patterns and potential relationships between the variables. Predictive data analysis is the process of using statistic or modeling techniques to make predictions about possible outcomes and can help people to make informed business decisions.

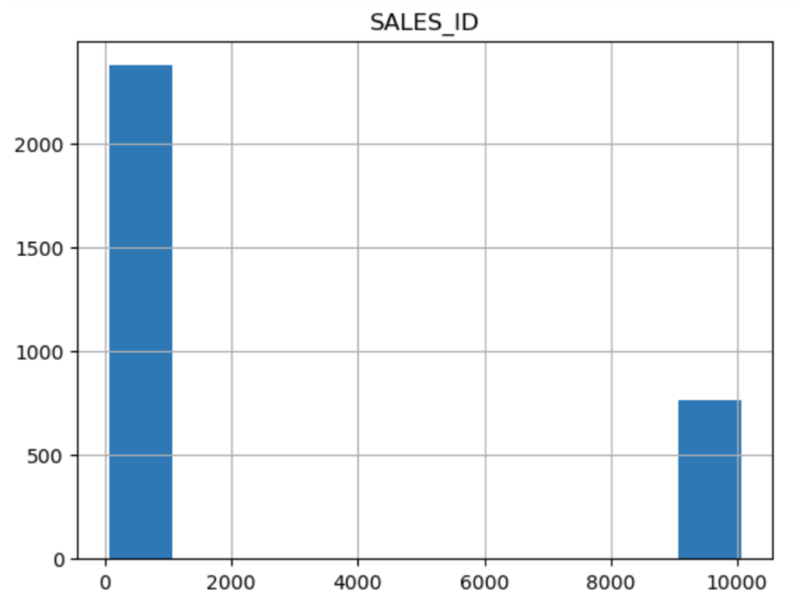

Fig. 3: SALES ID Histogram. This showed the distribution of the internal Realtor sales ID numbers. The large separation was known to occur because ids were created by two separate software packages.

3.3.1 Exploratory Data Analysis

Making exploratory data analysis an important part of the data analytics process allows for finding outliers/anomalies, testing hypotheses, and checking assumptions. Statistical summaries and different visualization techniques are used to summarize the data and present it in multiple ways, so the best understanding of the data can be developed. EDA is essential because it allows for evaluation of the data before assumptions are made. Identifying errors, outliers, and anomalies leads to a better understanding that can help form a more focused hypothesis: By performing EDA, better questions are asked, which then leads to a better focus, and thus a probable better outcome.. In essence, performing EDA moves towards a better model, which will most likely result in a more trustworthy, complete outcome.

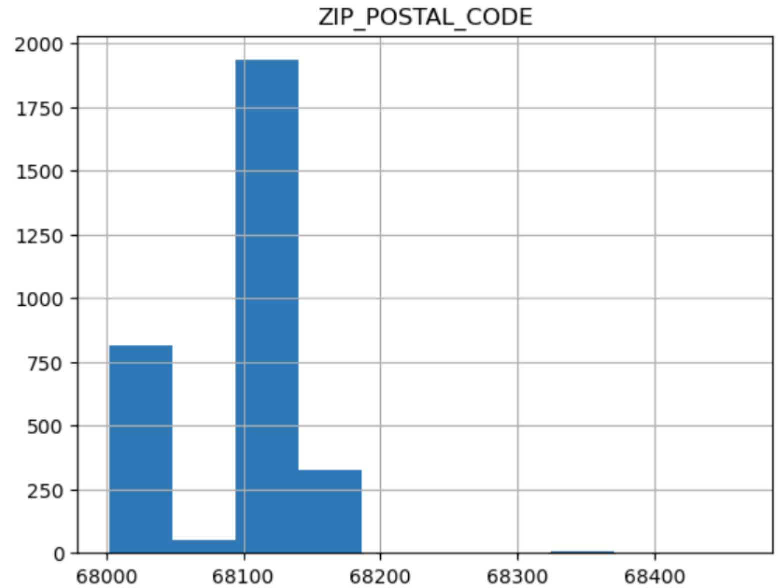

Fig. 4: ZIP/POSTAL CODE Histogram. This grouping demonstrated that all of the zip codes lied within the 68200 to 68200 range, which was known to represent the Omaha, NE, metropolitan area.

There are four main types of EDA:

1. Univariate non-graphical: This is a simple form of data analysis that is comprised of one variable. Because of this, there are no causal relationships, and the main purpose is to find patterns within the single variable.
2. Univariate graphical: Similar to the univariate non-graphical method, the univariate graphical method deals with just one variable. The difference, however, comes from the visualization of the patterns. (Most common: stem-and-leaf plots, histograms, and boxplots)
3. Multivariate non-graphical: A form of analysis that shows a relationship between 2+ variables. This analysis typically occurs through cross-tabulation or statistics.
4. Multivariate graphical: Like the univariate graphical analysis, the multivariate graphical analysis uses visualization to display relationships between the 2+ variables. (Most common: grouped bar plot, bar chart; other common: scatter plot, heat map, bubble chart) [14]

After cleaning the data, the CSV file was imported into a JupyterLab notebook, where the necessary Python modules (numpy, pandas, matplotlib, and seaborn) were then installed. The first step was checking the data to verify that the columns imported matched the CSV file using Python code. Python code was also used to preview the data in order to make sure it was displaying correctly. Next, Python was used to verify that no null values existed in the data. All main attributes of the data were verified as numerical in Python, which returned that it was dtype: int64. [3]

At this point, visualization was used to determine the distribution of the data (i.e., univariate graphical EDA). A command to visualize the OFFICE data in a boxplot was executed, and this visualization demonstrated that there were no outliers. (See Figure 2) Next, histograms were created in Python for both the SALES ID and ZIP/POSTAL CODE attributes to analyze the distribution of data. (See Figure 3 and Figure 4) Finally, the statistical information was reviewed within Python. [3]

Exploring multivariate visuals to look for patterns and causal relationships was the next step. A heat map was created to evaluate correlation values between the attributes. (See Figure 5)

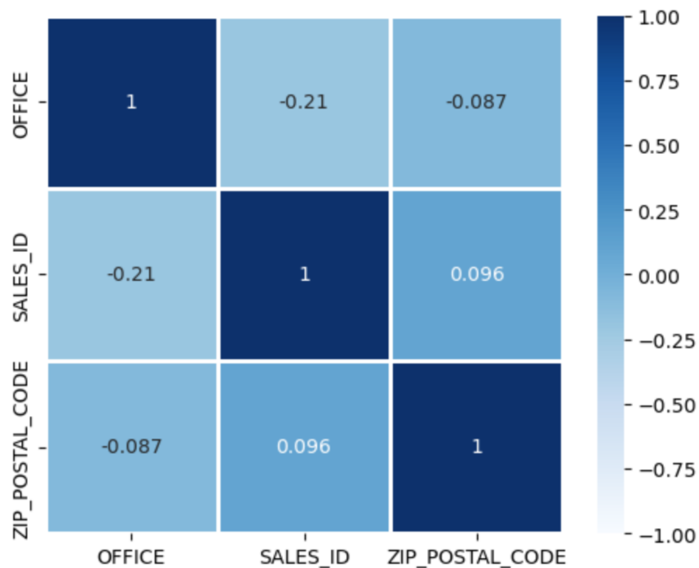

Fig. 5: Heat Map. This heat map showed the correlation values between the different attributes.

The heat map showed that the strongest correlation was a negative relationship between the SALES ID and OFFICE attributes (-0.21), while there was a minor positive relationship between the SALES ID and ZIP/POSTAL CODE attributes (0.096). [3]

Python was next used to create scatterplot visualizations to compare different attributes. The first scatterplot showed the distribution of transactions at each office location. As seen in Figure 6, the distributions appeared similar across each office.

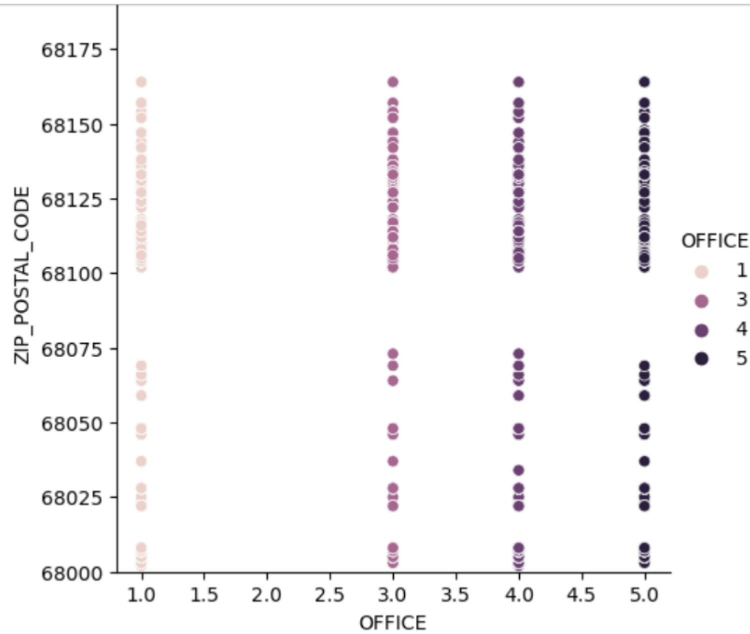

Fig. 6: OFFICE and ZIP/POSTAL CODE Scatterplot. This visualization showed the distribution of transactions' zip codes by office.

A second scatterplot showed the distribution of transactions by ZIP/POSTAL CODE for the SALES ID. (See Figure 7) This scatterplot was truncated after the SALES ID number 500 for better visualization. (This was due to the large split between SALES ID numbers, as seen in Figure 3.) [3]

At that point, scatter plots in Python were used to visualize the distribution of transactions across ZIP/POSTAL CODES by SALES ID and were executed for each unique office location. [3] This was an important step for reviewing the data, as it was more clear to see patterns in distribution (even minor ones) within each individual office. (See Figure 8)

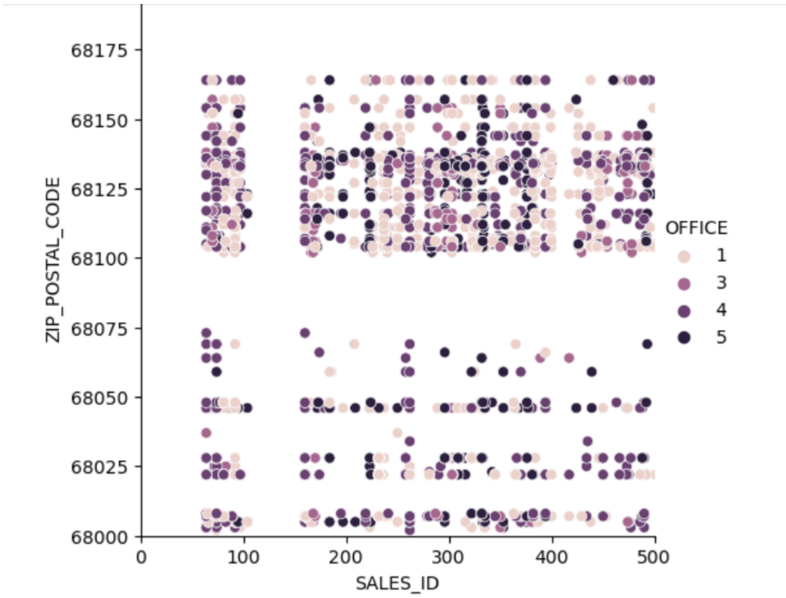

Fig. 7: SALES ID and ZIP/POSTAL CODE Scatterplot. This visualization showed the distribution of transactions' zip codes by sales ID number.

Each OFFICE was graphed this way using the following Python code:

Listing 1.1: Python Visualization Example

```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

# Visualization for distribution across each
individual office location

sns.relplot(x='SALES_ID', y='ZIP_POSTAL_CODE',
hue='OFFICE', data=data, row="OFFICE")
plt.ylim(68000, 68200)
plt.xlim(50, 500)
```

The Python notebook used for this project is available at this GitHub link.

Fig. 8: OFFICE 1 Only Scatterplot. This visualization showed the distribution of transactions' zip codes by SALES ID number in OFFICE 0001.

In order to create a geographical visualization of the distribution, the data was imported into Tableau. The data was then grouped and mapped by ZIP/-POSTAL CODE to show the number (by color-coded key) of transactions within each zip code. Each OFFICE location was added to the zip code for its location on the map to help demonstrate whether the office location impacted the surrounding area's sales numbers. (See Figure 9)

At this point, the Tableau map visualization allowed for the density of transactions to be viewed clearly: It became more noticeable that there were more transactions in the ZIP/POSTAL CODES surrounding OFFICE locations 0001, 0004, and 0005. This provided visual evidence that there were ZIP/POSTAL CODES that contained more transactions than others. Whether or not that showed a correlation between the OFFICE locations (which include the SALES IDs that operate in those locations) and the ZIP/POSTAL CODES of the transactions would be determined in the predictive analysis of the next phase in this project.

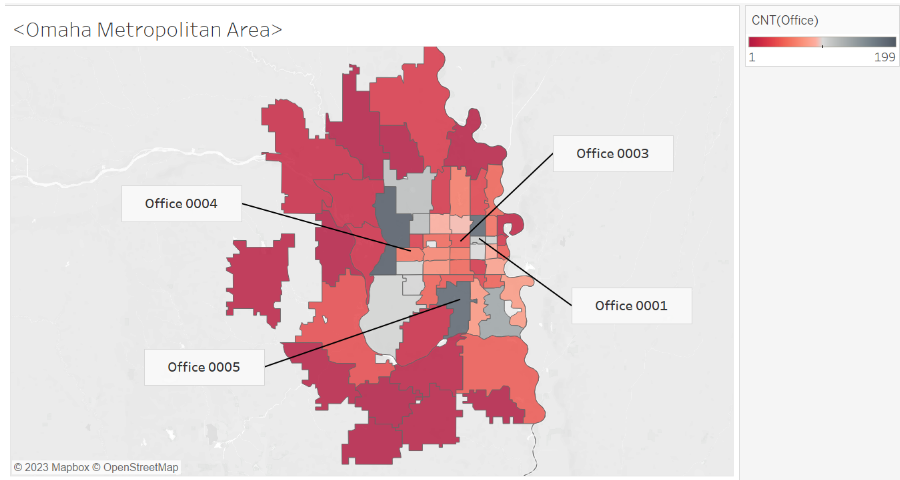

Fig. 9: Zip Code Map. This map showed the distribution of the transactions' zip codes on a color-coded map of the Omaha, NE, metropolitan . [4]

3.3.2 Predictive Analysis

In order to perform predictive analysis, a Machine Learning (ML) algorithm must be generated. This algorithm is built by following the steps of a Machine Learning Pipeline. The following steps describe how to build an ML pipeline:

1. Data collection - A data set is gathered from a verified source to be used to generate and support a hypothesis.
2. Data cleaning - The data set is processed and verified as good quality.
3. Feature extraction - The raw data is transformed into usable data that can be processed through an ML model, while still preserving the original data set.
4. Model validation - The transformed data set is split into train and test segments and then used to create and verify the model, as seen in Figure 10.
5. Visualization - Conclusions about the data set are made and published as a visual report. [11]

A Machine Learning algorithm is computer code that adapts to the data by recognizing patterns and then learns from it in order to produce a model that can be used to make predictions that help business decisions. For this project, the chosen ML algorithm needed to use the OFFICE location to evaluate the ZIP/POSTAL CODE. The ML algorithm chosen will have used the training data to create a model that will have aimed to predict whether a transaction's ZIP/POSTAL CODE was correlated to the OFFICE location.

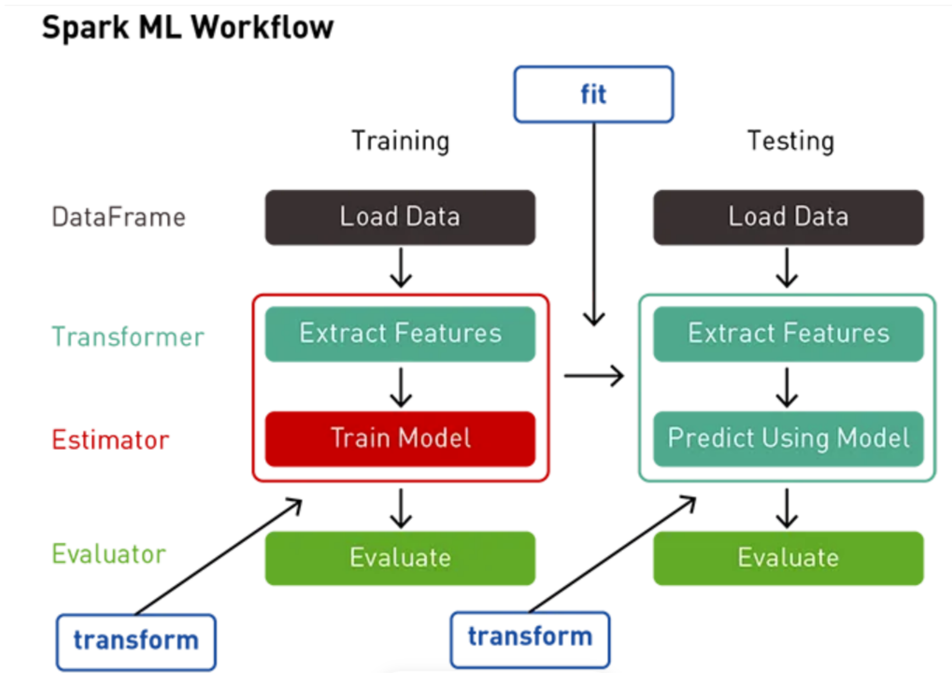

Fig. 10: A Machine Learning Pipeline. This visualization demonstrates the workflow of an ML pipeline. [11]

In Figure 11, the ZIP/POSTAL and SALES ID variables were used to determine the type of relationship between the variables; the OFFICE location for each transaction was added and was identified by the color key. The visualization of the data set clearly showed that there was no linear relationship between the variables, but instead that the data was grouped together. (The data can be found at this GitHub link [5].) Because of this, a Linear Regression algorithm was not suitable for this data set. Instead, classification algorithms were the appropriate choice, as they were most likely to produce a positive outcome from this data set. The classification algorithms used for this project were K Number Neighbors (KNN), K Means Clustering, and Random Forest.

3.3.2.1 Train/Test Data Set

To train and test the data, the Python module sklearn train/test split was imported to split the data into train and test segments. In order to use train/test split, the X and y variables first had to be defined; this allowed for the module to be used to create the Xtrain, ytrain, Xtest, and ytest variables. The following code showed the Python code used to create the train and test sets.

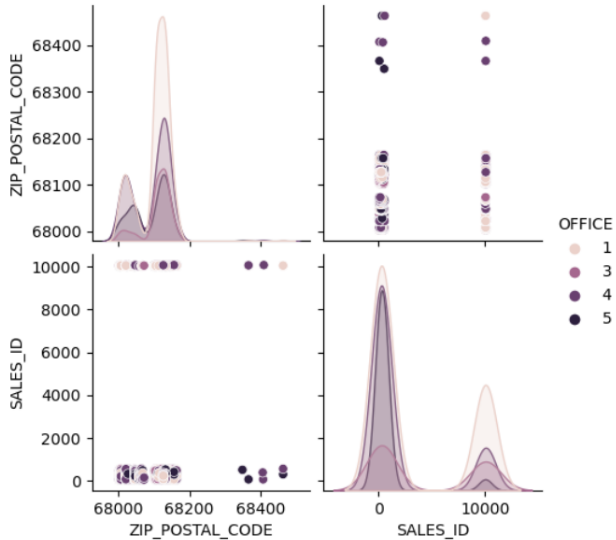

Fig. 11: Pairplot Visualization. A visualization of the data set (SALES ID and ZIP/POSTAL CODE), with the OFFICE identified by color. These graphs demonstrated the type of relationship that existed between the variables.

Listing 1.2: Python Train/Test Example

```
# Build train and test sets
from sklearn.model_selection import train_test_split
data_model = data[['OFFICE', 'ZIP_POSTAL_CODE',
'SALES_ID']]

# get dummy data
data_dum = pd.get_dummies(data_model)

# train test split
from sklearn.model_selection import train_test_split
X = data_dum.drop('OFFICE', axis =1)
y = data_dum.OFFICE
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
```

Next, three different sklearn modules (KNeighborsClassifier, KMeans, and RandomForestClassifier) were imported to train the data set and to create a model for future use. The training data was run through each sklearn module (each its own ML algorithm). Following this, the sklearn accuracy score module was imported and used to evaluate the accuracy of each model when the test data was used. The number of neighbors (KNN algorithm) and classifiers

(K Means algorithm) were tested with different constants (2-12) to determine the best accuracy score. Finally, three different measurements of success (error) were calculated on each model: mean absolute error (MAE), mean squared error (MSE), and confusion matrix.

The following code was from one of the sklearn modules, KNeighborsClassifier:

Listing 1.3: Python Module Example

```
# Train and test K Nearest Neighbor Model
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score

neigh = KNeighborsClassifier(n_neighbors=12)
neigh.fit(X_train, y_train)
y_pred_knn = neigh.predict(X_test)
accuracy = accuracy_score(y_test, y_pred_knn)
print("Accuracy:", accuracy)
```

Each of the modules followed a similar code structure. The full code can be found in the Python notebook used for this project and is available at this GitHub link.

3.3.2.2 Model Scores

The test data was processed by each model with an accuracy score having been determined immediately following the model's use. The final accuracy score was the best score attained after having tried different constants for the neighbors (KNN algorithm) and the classifiers (K Means algorithm) to determine the best possible accuracy score. The similar coding structure for each module allowed for the comparison of the process and the accuracy score to have occurred more smoothly.

Measurement scores were run for all three models to evaluate whether the model would be successful. The code for the measurement scores was run for each model individually, but they all used the same sklearn modules. The following Python code shows how the remaining three measurement scores for the KNeighborsClassifier model were found:

Listing 1.4: Python Measurement Example

```
# Additional measurement to for correlation - MAE
from sklearn.metrics import mean_absolute_error
mean_absolute_error(y_test,y_pred_knn)

# Additional measurement to for correlation - MSE
from sklearn.metrics import mean_squared_error
mean_squared_error(y_test,y_pred_knn)

# Another measurement for error - Confusion Matrix
```

```
from sklearn.metrics import classification_report,
confusion_matrix
pred_knn = neigh.predict(X_test)
print(classification_report(y_test,pred_knn))
```

The KNN model and the Random Forest model had very similar results when considering all measurement scores, although the Random Forest model did score slightly better than the KNN model in the Confusion Matrix. (See Table 1)

Table 1: Scores Used to Measure Models

Module	Accuracy	MAE	MSE	Confusion Matrix
KNeighborClassifier	0.624	0.92	2.646	0.63
K Means Cluster	0.203	1.891	5.971	0.38
Random Forest	0.638	0.912	2.692	0.65

The Random Forest model that was created was the best for predicting whether the OFFICE location and the ZIP/POSTAL CODE had any correlation. The results showed that there was a 65 percent chance that the OFFICE could be used to determine the ZIP/POSTAL CODE.

4 Conclusion

This project analyzed the potential relationship between real estate transactions and a Realtor's physical office location in order to examine whether it impacted real estate transactions. Real estate transactions from a local brokerage were used to compare the zip codes of a Realtor's transactions to that of their physical office location. The data set was evaluated with several algorithms and models were created to determine if a correlation existed. Based on the statistical evidence, the models showed 65% accuracy at best when trying to predict the correlation of office location with a zip code of a transaction. While this could be useful information in terms of possibilities for building office locations, more information would be needed in order to concretely recommend making that investment.

The GitHub repository for the entire project can be found at this GitHub link. [6]

4.1 Future Work

Future work could include extending this project to more than just the 2022 data set to determine if a larger time frame would show a higher accuracy of correlation. Perhaps the most intriguing and impactful future work for the owners of this local real estate brokerage, however, would be extending this project's philosophy to individual agents - to show whether their transactions are consolidated to specific areas (zip codes) within the Omaha, NE, metropolitan area. This information could be used to validate marketing strategies or suggest new target areas for marketing.

References

1. Agarwal, S.: (Feb 2018), https://www.sudeep.co/data-science/2018/02/09/Understanding-the-Data-Science-Lifecycle.html
2. Anderson, E.: 44688 capstone project / 2022 closed business journal csv data (Mar 2023), https://github.com/eden0214/44688_CapstoneProject/blob/main/2022_Closed_Business_Journal.csv
3. Anderson, E.: 44688 capstone project / anderson capstone python file (APR 2023), https://github.com/eden0214/44688_CapstoneProject/blob/main/Anderson_Capstone.ipynb
4. Anderson, E.: 44688 capstone project / zip code map (Apr 2023), https://github.com/eden0214/44688_CapstoneProject/blob/main/Zip_Code_Map.png
5. Anderson, E.: 44688 capstoneproject 2022 closed business journal cleansed.csv at main · eden0214 / 44688 capstoneproject (Apr 2023), https://github.com/eden0214/44688_CapstoneProject/blob/main/2022_Closed_Business_Journal_CLEANSED.csv
6. Anderson, E.: Eden0214 / 44688 capstone project (Mar 2023), https://github.com/eden0214/44688_CapstoneProject
7. Anderson, J., Bandur, M.: How 'bidding wars' are dramatically impacting valuations for omaha homeowners (Mar 2022), https://www.ketv.com/article/omaha-homeowners-valuations-real-estate-market/39282367
8. Barwick, P.J., Wong, M.: Competition in the real estate brokerage industry: A ... - brookings (Dec 2019), https://www.brookings.edu/wp-content/uploads/2019/12/ES-12.12.19-Barwick-Wong.pdf
9. Florida, R.: What makes us the happiest about the places we live (Sep 2014), https://www.bloomberg.com/news/articles/2014-09-19/what-makes-us-the-happiest-about-the-places-we-live
10. Görener, A., Hacıoglu, U., Dincer, H.: (Apr 2013), https://www.ssbfnet.com/ojs/index.php/ijfbs/article/view/430/385
11. M, S.: What is a pipeline in machine learning? how to create one? (Dec 2019), https://medium.com/analytics-vidhya/what-is-a-pipeline-in-machine-learning-how-to-create-one-bda91d0ceaca
12. 7 reasons to work with a realtor® (Jun 2011), https://www.nar.realtor/magazine/tools/client-education/handouts-for-buyers/7-reasons-to-work-with-a-realtor
13. Newell, S.J., Plank, R.E.: The importance of relationship and consultative ... - baylor university (Mar 2012), https://www.baylor.edu/business/kellercenter/doc.php/194347.pdf
14. Newsletter, I.: What is exploratory data analysis?, https://www.ibm.com/topics/exploratory-data-analysis
15. The omaha metro (Oct 2020), https://rivercityomaha.com/the-omaha-metro/
16. Weinstock, L.R.: Introduction to u.s. economy: Housing market (Jan 2023), https://sgp.fas.org/crs/misc/IF11327.pdf