

## Public Sentiments on Vancouver vs Calgary

Comparative Sentimental Text Analysis

LING 450: Computational Linguistics

Prof. Maite Taboada

Name	Student ID	Contribution
Monisha Mohandas	301384925	<ul style="list-style-type: none"><li>• Sentiment Analysis using VADER</li><li>• Sentiment Analysis using Manual Annotations</li><li>• Methodology using VADER</li><li>• Case Study</li><li>• Overall editing, formatting and support</li></ul>
Olivia Calton	301291735	<ul style="list-style-type: none"><li>• Sentiment Analysis using Manual Annotations</li><li>• Manual Annotations for 6 Texts</li><li>• Introduction, Challenges and Limitations</li><li>• Methodology using Manual Annotation Process</li><li>• Overall editing, formatting and support</li></ul>
Mithula Barua	301217552	<ul style="list-style-type: none"><li>• Sentiment Analysis using Manual Annotations</li><li>• Manual Annotations for 6 Texts</li><li>• Methodology using Manual Annotation Process</li><li>• Overall editing, formatting and support</li></ul>

## Introduction

Recently, there has been an increase in British Columbia's provincial emigration rates. "Over the last year, 67,944 Canadians left B.C. Statistics Canada says the majority of them — about 37,650 people — packed up their belongings and moved to Alberta," (Ballard, 2024). Furthermore, we found that most Canadians, from each province, are moving to Alberta (Alberta Net Migration, 2024). We wanted to learn further about what factors are causing this, and overall public sentiments on the issue.

There were many reasons for comparing the results between Calgary and Vancouver. The immigration rates to Alberta were certainly ones to consider, however, our research also relied on the stark differences in the two cities to help us see how they compare. Some of the reasons such as the cities geographies, locations, weather, urban vs rural settings, social demographics, and most significantly home prices, all played a role in us choosing Calgary as a comparative city to Vancouver.

We wanted to see if we can find any basis on how quality of life is different within the two cities, and whether the stark housing prices in Vancouver affect individuals' quality of life, as opposed to in Calgary, which proposes some of Canada's lowest housing prices. In order to mitigate some biases, we decided to poll a larger sample size. Furthermore, we looked at other sites such as Quora, finding that online sentiments remained similar across different websites.

## Methodology

Our analysis commenced with the extraction of texts from pertinent Reddit posts. Our corpus consisted of 20 reddit posts in total from comments ranging from 50 to 200, and these posts were segmented into three distinct files to serve specific research objectives:

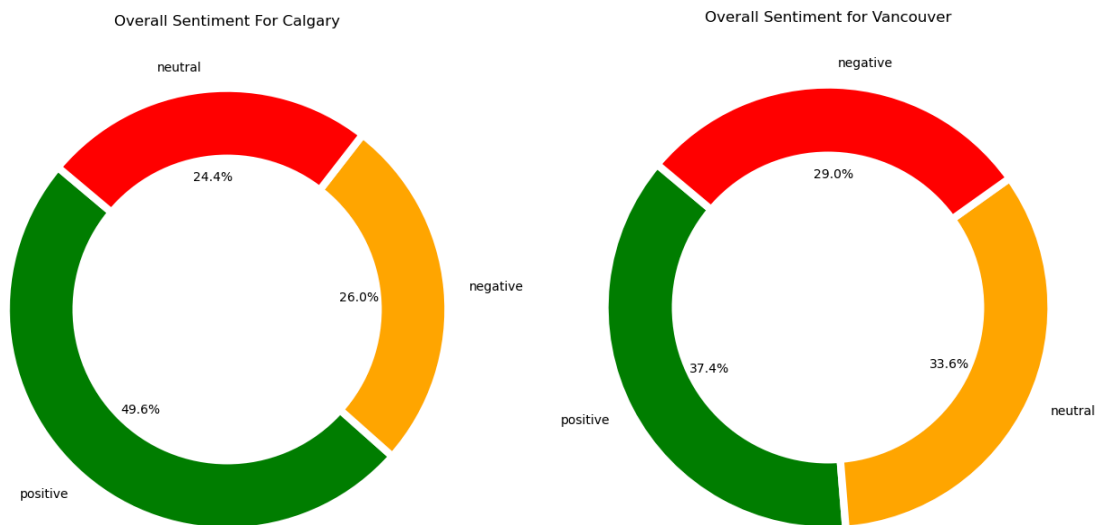
1. The first file collated threads on housing and affordability in Vancouver to gauge the impact of the housing crisis on the residents and the severity thereof.
2. The second file encompassed discussions from individuals who relocated to Calgary from other provinces, aimed at understanding the factors that make Calgary an attractive destination, particularly in light of observed interprovincial migration trends.
3. The third file aggregated dialogues from those who moved from Vancouver to Calgary, honing in on the reasons behind this specific migratory pattern out of British Columbia.

Following data collection, we conducted sentiment analysis on each corpus individually. With the aid of GPT models, we filtered out advertisements and usernames,

retaining only content pertinent to sentiment evaluation. The cleansed texts were then processed in Jupyter to remove extraneous whitespaces and lines, ensuring data integrity. These texts were then structured into a single-column dataframe and subjected to sentiment analysis using the VADER Sentiment Intensity Analyzer.

The resulting sentiment distribution was as follows:

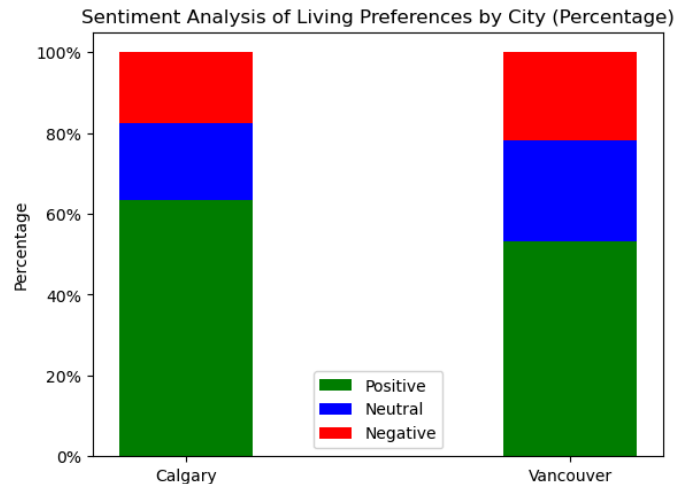
Sentiment	Calgary	Vancouver	VancouverVSCalgary
Positive	390	304	213
Neutral	192	273	112
Negative	205	236	75



Key insights from this analysis indicated that Calgary's residents generally expressed higher satisfaction, largely attributed to a lower cost of living and housing affordability. In contrast, Vancouverites expressed significant frustration with exorbitant rents, with a notable trend of relocating to more financially lenient provinces. However, in threads directly comparing the two cities, users extolled Vancouver for its mild climate, diverse culture, and culinary scene, particularly highlighting the sushi offerings. Despite these positives, the prevailing sentiment underscored substantial lifestyle adjustments or relocations as a response to Vancouver's high living costs.

To further delve into the comparative sentiment of both cities, we pivoted from a general sentiment analysis to a keyword-centric approach. We labeled comments with

city-descriptive keywords, followed by a targeted sentiment analysis, which yielded more granular insights into the cities' perception. This allowed for a direct sentiment comparison within the same corpus.



Additionally, we tokenized the text and identified the top 150 words from each corpus to generate corresponding word clouds.

## Most Common Words and Word Clouds for Vancouver VS Calgary posts



Analyzing common words from Reddit discussions about relocating from Vancouver to Calgary reveals a rich tapestry of considerations that users weigh. Geographic identifiers like 'Calgary,' 'Vancouver,' 'Alberta,' and 'BC' dominate the

conversation, indicating that the debate is deeply rooted in the nuances of place. Discussions often pivot on comparative evaluations, with terms such as 'better,' 'expensive,' 'cheaper,' 'affordable,' and 'quality' reflecting the cost and quality of living in each city. Lifestyle elements, as signified by 'school,' 'food,' 'restaurants,' 'mountains,' and 'activities,' suggest that cultural and recreational opportunities are pivotal to these deliberations. Sentiments within the threads vary, with positive expressions like 'like,' 'great,' 'good,' 'love,' and 'beautiful' intermingled with more negative terms such as 'bad,' 'expensive,' and 'miss,' portraying a complex emotional landscape. The significant frequency of 'rain,' 'cold,' 'weather,' 'winter,' and 'sunshine' points to climate as a key factor influencing relocation decisions. 'Housing,' 'rent,' 'prices,' and 'afford' consistently emerge, emphasizing housing affordability as a central theme. The importance of social factors is also evident through words like 'family,' 'kids,' 'friends,' and 'community,' underscoring the value placed on social ties and community cohesion. Practical daily considerations are not overlooked, as evidenced by mentions of 'transit,' 'drive,' 'traffic,' and 'commute,' indicating that the logistics of transportation are also a significant factor in the moving dialogue. This analysis not only showcases the multifaceted considerations discussed by potential movers but also provides invaluable insights into the collective mindset regarding what constitutes a satisfactory urban living experience.

### **Overcoming Data Collection Hurdles**

The data collection phase presented distinct challenges, particularly the scarcity of Calgary-centric Reddit posts explicitly discussing housing and affordability. This was in contrast to the volume of Vancouver-related threads, likely indicative of the pronounced housing crisis and its salience as a topic of public discourse. Despite this, the Calgary threads were informative; they frequently cited lower cost of living and housing expenses as significant draws to the city, occasionally complementing these economic advantages with references to a preferable climate when compared to cities like Toronto.

### **Discrepancies in Sentiment Analysis**

Preliminary manual review of Calgary's Reddit discussions suggested a prevailing positive sentiment, conflicting with the significant negative sentiment in VADER's overall sentiment analysis. Conversely, Vancouver's threads predominantly reflected a negative sentiment, aligning with the city's ongoing housing predicament. However, the VADER sentiment analysis presented a significant amount of positive sentiment which called for further examination.

This disparity between anticipated and VADER derived sentiments highlighted the imperative for manual annotation. The necessity for this granular approach was twofold: to authenticate the veracity of the automated sentiment analysis and to capture

the nuanced expressions of sentiment often lost in automated processes. The subsequent manual annotation was thus integral to our methodology, enhancing the reliability of our sentiment assessments.

This meticulous manual examination allowed for a deeper comprehension of the sentiments expressed by Reddit users, informing a more nuanced interpretation of public opinion. It also served to validate the automated sentiment analysis, ensuring the academic integrity and robustness of our research findings. The specifics of this manual annotation process and its implications for our study are detailed in the ensuing section of our paper.

## **Manual Annotation Process**

We began the process by first annotating 12 out of our 20 texts. Quickly, we raked up 84 pages, and realized we had too much data for manual annotations, a common issue with manual annotations being how difficult and time consuming they can be. In order to compare our manual data with VADER code, we picked 5 random comments from our 20 texts, in order to compare identical datasets. For Manual Annotations, we each made comments on whether we perceive a certain text to be positive, negative or neutral, and found a 96% consensus in our findings. However, VADER only agreed with us 25%, a striking difference in the two methodologies.

For our Manual Annotation process we divided the section into several steps to ensure that the codes have picked up the right sentiments for all 88 comments. We began the process by individually processing the comments with -1, 0, and +1. We read the comments and categorized the overall sentiments of people based on their comment. The categorization consists of positive (+1,) negative (-1), and neutral (0). After identifying the categorization of 5 comments each from all 20 text files we used, we compared VADER's output to our scoring of the sentiments. Certain words in the sentence may indicate negative sentiments but the whole comments could still be positive because the context of the comment suggests that it is positive rather than negative. Therefore, the context of the annotation was important in categorizing whether the comments should be negative, positive, or neutral.

We inserted all 88 comments from the five selected comments of each text file from Calgary and Vancouver into an excel file and have the VADER score and our score along the row to indicate whether the text analysis from the code was able to correctly figure out the sentiments. A very significant conclusion based on the result is that the VADER rating accuracy was at 25.00%. The rate is significantly lower than our scoring suggesting that VADER agreed with our scores only 25% of the time which means that

VADER is not good in understanding the context or sarcasm or negative words used in positive sentences and so on.

The presence of sarcasm makes it difficult for VADER to detect any form of sarcasm or ambiguities. The code does not have the sufficient understanding in detecting sarcasm which is the reason for the low rating compared to our manual annotation process. Some examples where the VADER code did not pick up the right sentiment because of sarcasm is shown in comment 22 “Even Calgarians don't realize how good winter is here. The VADER selected this comment as negative but in fact it should be positive because the person is expressing how Calgary winter is better which has no negativement sentiment. Another example where the VADER incorrectly identified the sentiment is found in comment 63 which states that “Don’t recommend moving here. Not only is it extremely expensive, lots of people are having a challenge finding sustainable work here as well (my friend couldn’t find work for 5 months) and you’ll want to make over \$25 an hour to afford a cheaper rental(lower end for a studio or 1br is anywhere between \$1500-\$2000). If you’re set on moving this way, find roommates, otherwise it can be very challenging here“. The VADER code identified this comment as positive but should have been negative. The comment is addressing the negative sentiment about the housing cost in vancouver and the person does not like it. Another example is found in comment 77 stating “I have to continue living with my ex because I can't afford to move out“. The VADER code identified this comment as neutral but this should have been negative. The person is addressing negatively how he needs to stay with girlfriend to afford living in Vancouver. Based on these three examples, we can see that VADER can incorrectly assess sentiments.

### **Insights From Manual Annotation Process**

We learned that VADER is not an efficient method for conducting text analysis as compared to our manual analysis where we had a 96% consensus between the three of us, VADER’s result was only 25%.

Another key insight we found during our text analysis was how much context matters because that was a critical aspect needed to fully understand the expressed emotions by people. This is because we found out that oftentimes code does not account for context as humans can. Moreover, the presence of sarcasm makes it difficult for VADER to detect any form of sarcasm. This is because the code lacks the guidance to assess irony, sarcasm and ambiguities and cannot assess between positive and negative. We found out that oftentimes code does not account for context as humans can. For example, it also does not manually weigh the headings of a certain text as heavier, etc. For us, it didn’t understand that the headings were different, and counted them as texts.

## **Challenges And Limitations**

The limitation of our research is, we recognize that the sentiments off of Reddit are often strong, often strongly positive or negative, and fewer neutral viewpoints. However, we deemed this would be the best place for us to really be able to view the strong sentiments that motivate individuals and their decisions to migrate to and from different provinces.

The scope of our project is sentiment analysis from Reddit, regarding public opinions on Calgary vs Vancouver. Our project does not provide an accurate view on perspectives on other cities, or overall public opinion, simply those represented on the Reddit forum. We also tried to pick a broad scope of reddit posts that captured different aspects of living in the city including governance, rent, lifestyle factors, and more.

## **Case Study**

### **Twitter Sentiment Analysis: A Case Study for Apparel Brands**

Authors: Abdur Rasool, Ran Tao, Kamyab Marjan, Tayyab Naveed

Published: Journal of Physics: Conference Series, 2019

The proliferation of social media has significantly influenced consumer behavior and brand perception. This study focuses on the sentiment analysis of tweets related to two major international apparel brands: Adidas and Nike. By analyzing Twitter data, the study aims to determine public sentiment towards these brands and assess how online discussions can inform marketing strategies.

## **Methodology**

The study utilized data scraped from Twitter, employing Tweepy to capture tweets that mentioned Adidas and Nike. A total of 99,850 tweets were collected, preprocessed to remove non-English content, duplicates, and noise such as URLs and redundant characters. The cleaned dataset comprised 17,006 tweets, which were analyzed using the Naïve Bayes algorithm and lexicon dictionaries to classify sentiments as positive, neutral, or negative.

Sentiment analysis revealed that Adidas received more positive mentions compared to Nike, although the difference in negative reviews between the two brands was marginal. The study highlighted that Twitter users frequently compare brands when discussing their preferences, indicating a competitive landscape even within individual tweets.



This research underscores the critical role of social media in shaping brand perception and consumer decision-making. The sentiment analysis not only provided insights into consumer preferences but also revealed broader trends in online brand engagement. For instance, the data showed a significant number of neutral sentiments, suggesting that while many tweets did not express strong opinions, the mere mention of the brands contributes to their visibility and engagement metrics.

## **Conclusion**

The case study demonstrated that sentiment analysis could yield valuable insights into public opinion on social media, guiding brand strategies and marketing decisions. Adidas was found to have a slightly better positive sentiment score than Nike, which could influence future marketing strategies. The study also proposed the potential for more refined analyses, incorporating a broader range of apparel brands and enhanced analytical techniques to improve accuracy and relevance.

## **Future Work**

Further research is suggested to include a larger dataset of multiple brands and to develop a domain-specific sentiment analysis tool that better captures the nuances of apparel-related discussions on social media.

## **Summary of SDA490 Project**

Our project delved into the intersection of the housing crisis and quality of life across Canada, particularly focusing on the varying factors that contribute to life satisfaction. We looked into data from Statistics Canada, surveys, public opinion through reddit, and Canadian Mortgage Housing Corporation (CMHC), which was used to conduct data analysis using R. It's noteworthy that, despite the housing challenges, a majority of Canadians report a high degree of life satisfaction. Factors such as race, sexual orientation, and diverse abilities were all found to influence the quality of life. Furthermore, it was evident that shorter commutes and better access to amenities and services significantly enhanced life quality. Yet, there is a palpable discontent with the housing prices in Vancouver, compelling some residents to consider relocating or moving out of the city. This discontent was most acutely felt by specific groups such as renters, students, and immigrants, who are more adversely impacted by these inflated costs. In conclusion, our findings underscore a robust correlation between exorbitant housing prices and the diminishment of quality of life. High costs of living leave individuals with less disposable income and lower overall affordability, which in turn limits their ability to fully enjoy life. This underlines the critical nature of our investigation into the pervasive issue of housing affordability and its direct bearing on the well-being and satisfaction of individuals in Canadian cities.

## References

Alberta Treasury Board and Finance. (n.d.). Net migration. Economic Dashboard.

Retrieved from <https://economicdashboard.alberta.ca/dashboard/net-migration/>

CBC News. (2023). B.C. sees 70,000 people exodus in 2023 as many pack up their

belongings and move to Alberta. CBC. Retrieved from

<https://www.cbc.ca/news/canada/british-columbia/bc-70-thousand-people-exodus-1.7159382>

Rasool, A., Tao, R., Kamyab, M., & Naveed, T. (2019). Twitter Sentiment Analysis: A

case study for apparel brands. *Journal of Physics. Conference Series*, 1176,

022015. <https://doi.org/10.1088/1742-6596/1176/2/022015>