Final Capstone Project

A Data-Driven Guide to Berlin's Neighborhoods for University Students

by Thithat (Theo) Promlikitchai

1. Introduction

Berlin is one of the most populous cities in the European Union, with an estimated population of 3.6 million in 2021. The city's rich history of immigrants and its inclusive attitude towards diverse identities have made the city an international destination for highly skilled, young talents. Besides its cultural assets and lively social scenes, Berlin also boasts several world-renowned universities and research institutions recognized by the Times Higher Education such as the Free University of Berlin and the Humboldt University of Berlin. Furthermore, Berlin offers a great selection of post-graduation opportunities in its start-up world or one of its fast-growing service, creative, and high-tech industries.

This project aims to respond to Berlin's ability to attract both domestic and international students by tackling one of the biggest concerns facing this demographic: relocating to and settling in a new city. Thus, the problem statement can be phrased as the following questions:

> *Which neighborhoods in Berlin are most appealing to university students? How can we characterize the experiences of living in one of these neighborhoods?*

Although several subjective factors will contribute to such a decision, this project will examine quantifiable factors such as average rent prices, demography, and average commute distances to surrounding universities. In addition, venues surrounding public transport stations will be used to create different clusters. These clusters will help describe the amenities available in each neighborhood and further tailor the project's recommendations to the unique needs of each student.

1.1 Business Case

This project will benefit stakeholders including but not limited to:

1) *Prospective University Students:* Students will be able to narrow down their apartment search to areas that best suit their lifestyle choices and other specifications, making the process of resettling more convenient.
2) *The City of Berlin:* The municipal government can use data science to help brand Berlin as a student city and promote specific neighborhoods, bringing in additional revenue and an educated workforce.
3) *Real Estate Companies:* Commercial and residential real estate companies that target university students and young adults can leverage such findings to support their investment decisions and marketing strategy.

2. Data

Below is a list of data sources used in this project and their description:

1) Berlin's localities GeoJSON file for choropleth mapping and CSV file with names, the boroughs they belong to, and respective land areas
   - Berlin is divided into 12 boroughs (Bezirke), which are then subdivided into a total of 96 localities (Ortsteile).

- Additional information can be found and webscraped from Wikipedia.

2) A CSV file with a list of universities (Hochschulen and Universitäten) in Berlin obtained from HochschulKompass and Wikipedia
   - Coordinates were manually coded using Google maps.
3) Public transport stations used for clustering
   - The raw Excel file was cleaned by referencing Wikipedia to include only U-Bahn (metro) and S-Bahn (light rail) stations.
   - Stations beyond the boundaries of Berlin and duplicated stations were dropped, as well as bus & tram stations since the data for these two categories were incomplete.
   - Coordinates were then converted from the Gauß-Krüger Zone 4 to WGS84 Decimal Degrees.
   - Rather than using localities, clustering these stations will provide more variation in the results. Furthermore, students who relocated to an urban area are unlikely to own a car, making access to public transportation essential.
4) A CSV file containing demographics by localities
   - Data as of December 31st, 2019. The raw CSV file was cleaned by referencing the accompanying metadata.
5) Rent prices by localities webscraped from HomeDay
   - Data as of 2019. Units is in €/m^2.
6) Foursquare venue data obtained using the Foursquare API
   - In order to segment the public transport stations into clusters, venues within a 700 meters search radius of each station were collected. For this project, neighborhoods are not defined by the localities but rather by the walking catchment around these stations.
   - 700 meters was selected based on the average walking distance of Berliners, and this distance is unlikely to result in significant overlapping judging from the average walking distance between stations.
   - Foursquare metadata was also downloaded to group venue subcategories during data processing.

The Python libraries used in this project include:

- *Webscraping:* BeautifulSoup, requests, json, os
- *Data Processing:* numpy, pandas, geopandas
- *Data Visualization:* matplotlib, seaborn, folium
- *Spatial Data:* geopy, pyproj, haversine
- *Machine Learning:* sklearn
- *Miscellaneous:* warnings, tqdm, IPython, dataframe_image

3. Methodology

3.1 Creating the Berlin Localities dataframe

The CSV file containing Berlin's 96 localities was read into a dataframe and cleaned using pandas. After verifying that the number of boroughs and localities in the dataframe totaled to 12 and 96 respectively, the geographical coordinates of each locality were retrieved using geopy's geocoder. After comparing the geocoded latitudes and longitudes with the GeoJSON file, it was found that two localities were incorrectly coded and were subsequently recoded. These two localities were

Reinickendorf and Pankow, which were assigned coordinates of their boroughs that happen to have the same name.

Below is a [sample map of Berlin]() obtained from Google:

Figure 1: Map of Berlin

3.2 Creating the Berlin Universities dataframe

After obtaining the names of 39 universities from HochschulKompass and Wikipedia, their coordinates were manually coded using Google Maps, and the resulting CSV file was compiled using Excel. Geopy's geocoder wasn't able to code all of the universities in the list, possibly because some of the university names have special characters that may not have been translated correctly by the API.

The distribution of universities in Berlin was then visualized using Folium. For every Folium map in this project, the GeoJSON file outlining Berlin's localities was read using geopandas and added onto the map as the background to provide labels and the boundaries for each locality.

As expected, universities in Berlin tend to be concentrated in the city center, while a few outliers are located in the outer localities, which is evident in Figure 2.

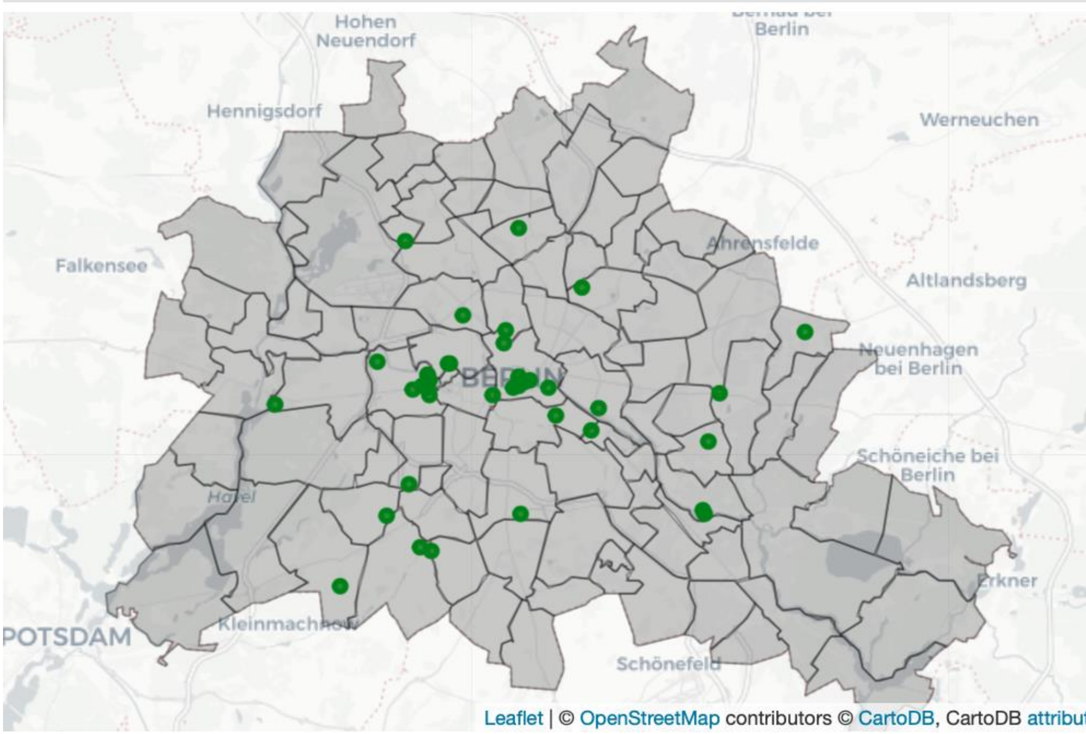


Figure 2: Map of Universities in Berlin

3.3 Creating the Train Stations dataframe

As mentioned in the Data section of the report, in this project, neighborhoods were defined as the 700 meters walking catchment around S- and U-Bahn stations. Thus, in order to create a dataframe that is ready for mapping and the Foursquare API, the raw Excel file was read using pandas. Only building structures were selected for further processing since the dataset also contained data points for entrances, transfer points, and platforms.

The coordinates were formatted as Gauß-Krüger Zone 4, which needed to be converted to WGS84 Decimal Degrees using pyproj. After transforming the coordinate system, stations that were beyond the boundaries of Berlin, duplicated stations (i.e., an S- or U-Bahn station that was double-coded as S+U), and bus & tram stops were dropped. These remaining 277 stations were then reverse coded to obtain the locality and borough that each station belongs to. Finally, the distribution of S- and U-Bahn stations in Berlin was visualized using Folium.

Based on Figure 3, not every locality in Berlin contains an S- or U-Bahn station. As expected, stations also tend to be more concentrated near the city center, but it is important to note that the outer localities are more likely to be serviced by bus or tram lines, rather than having a non-existent public transportation system.

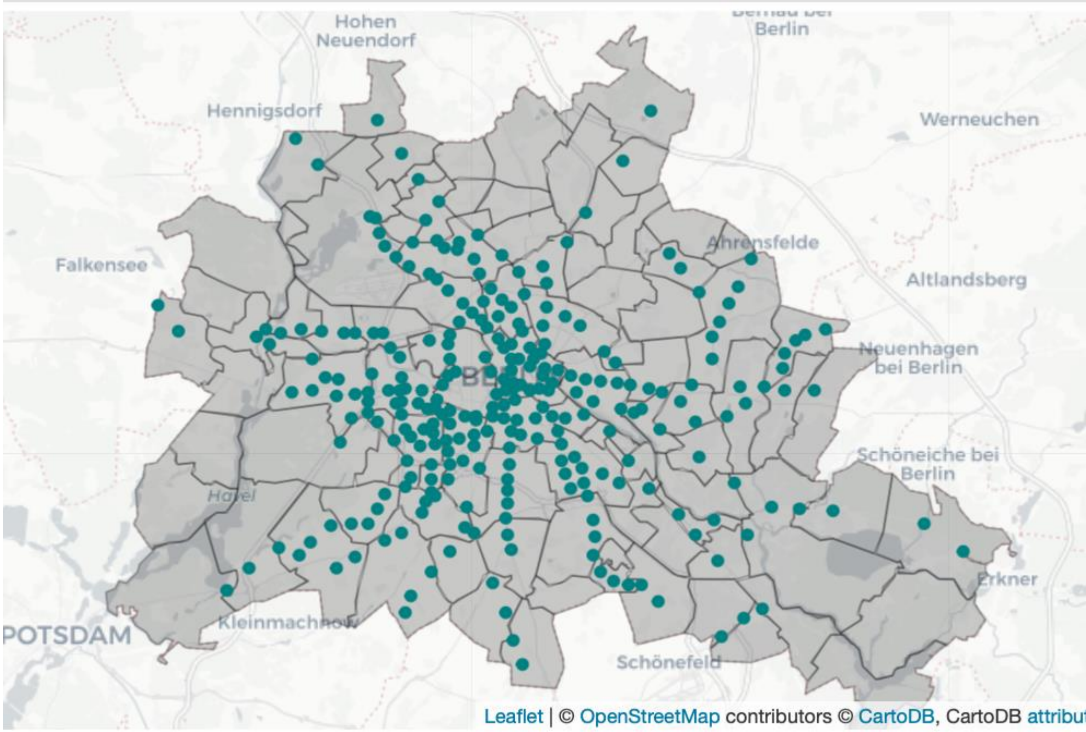


Figure 3: Map of S- and U-Bahn Stations in Berlin

3.4 Creating the Demographics dataframe

The first criterion used to determine the final list of suitable localities for university students was demography. The raw CSV file containing demographic information by localities was downloaded and read into a dataframe with pandas. Referencing the metadata, gender, citizenship, and age groups were cleaned and reorganized. Moreover, 15- to 30-year-olds were designated as young adults and the proportion of young adults to the total population of each locality was calculated and visualize via Folium as a choropleth map. Lastly, localities with the highest proportion of young adults, or the top 50%, were considered to have passed the first criterion and were subsetted.

Judging from Figure 4, localities in central berlin tend to have a higher proportion of young adults compared to localities on the periphery. However, confirming whether a statistically significant spatial autocolerration exists would have to be done using a geographic information system.

3.5 Creating the Rent Prices Dataframe

The second criterion was housing expenses. Rent prices were webscraped from HomeDay using BeautifulSoup and processed into a dataframe using pandas. 2019 rent prices were then visualized via Folium as a choropleth map. Finally, localities with extremely high rent prices (top 20%) were removed from the dataset.

Based on Figure 5, the most expensive localities are unsurprisingly the most popular places to live in Berlin such as Tiergarten, Friedrichshain, Kreuzberg, and Prenzlauer Berg, while the cheaper localities are located on the outer edges of Berlin, especially the Eastern side.

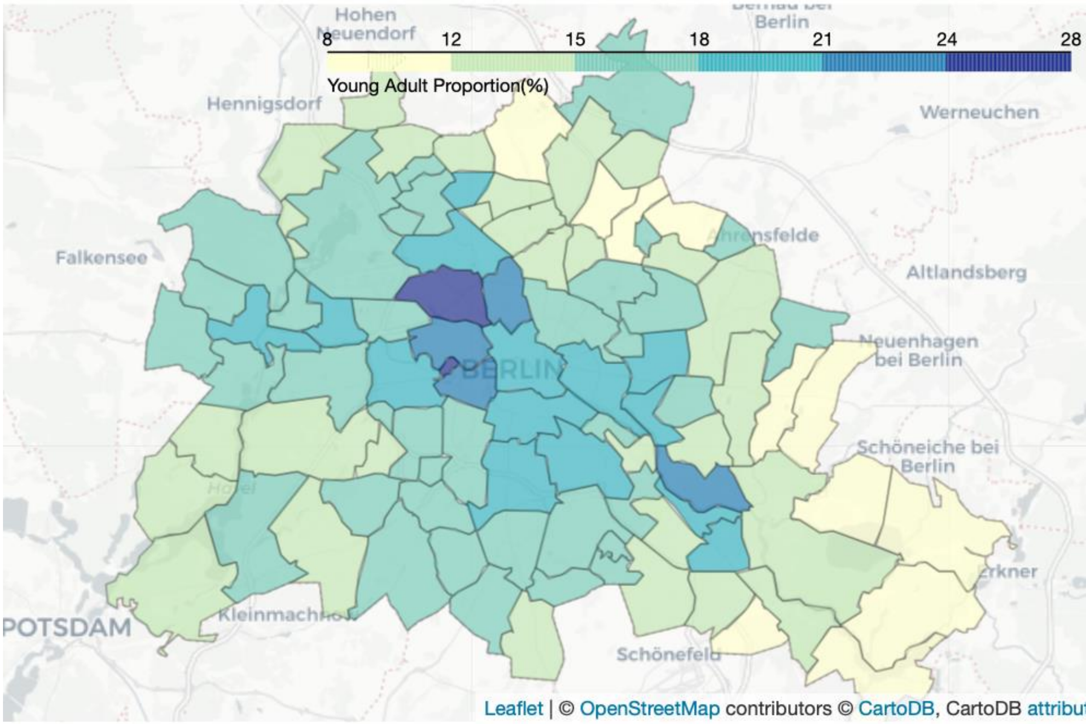


Figure 4: Proportion of Young Adults by Locality

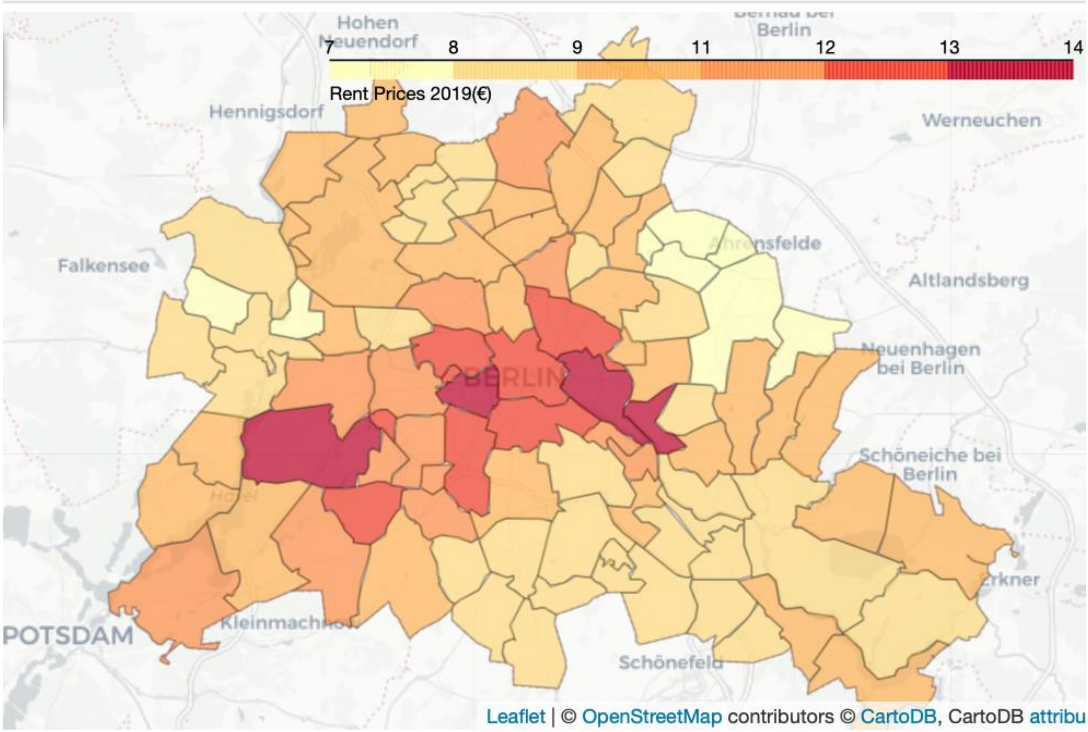


Figure 5: 2019 Rent Prices by Locality

3.6 Calculating Distances from Universities to Localities

The final criterion to consider was the commute distances. Students would naturally prefer to live in an area where commuting to their university is convenient and quick. The haversine distance from each locality to each university was calculated and the results were compiled into a dataframe using pandas. Keeping the coordinates of each locality and university for every pairing allowed for the names to be retrieved using merge operations.

According to Moovit Insights, the average commute distance for Berliners on a one-way trip is 8.36 kilometers. Using this value as the reasonable commute distance, the number of universities one would be able to reach from each locality was calculated, and the localities with the highest count (top 50%) were subsetted.

3.7 Selecting Localities for Cluster Analysis

By comparing the 48 localities with the highest proportion of young adults, the 77 localities with acceptable rent prices, and the 48 localities that are closest to universities, a final list of 20 localities that passed all three criteria was created. This list was then used to select S- and U-Bahn stations that will serve as origin points when obtaining venue data for clustering.

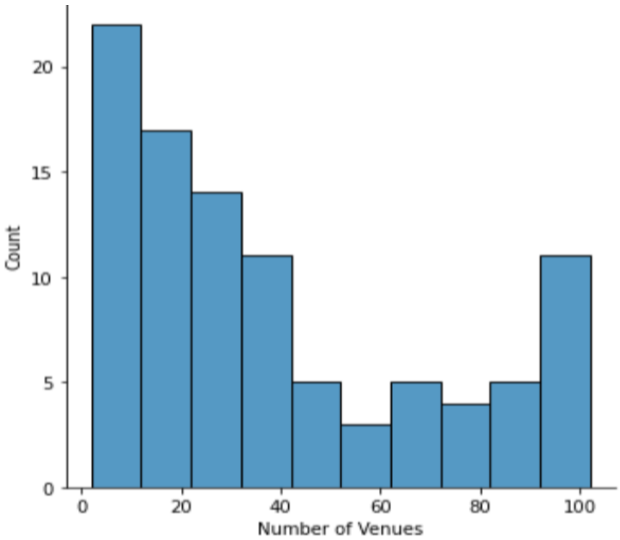
3.8 Obtaining and Cleaning Venue Data using the Foursquare API

Using the 97 stations mentioned in the step above, all venues within a 700 meters radius of each station was collected via the Foursquare API and read into a dataframe using pandas. Out of the 3,963 venues retrieved, 195 venues were dropped since venues that are transportation-related or are outdoor objects are unlikely to provide much information on the amenities available in each neighborhood.

Before proceeding to clustering, Foursquare's metadata on venue categories was downloaded and exploratory data analysis was performed. Foursquare has a total of 10 main categories and 3 additional tiers of subcategories. The original dataframe contained the name of each venue, its coordinates, and subcategory; adding on each venue's main category would help facilitate grouping operations. The frequency of each main venue category was then determined as seen in Table 1, and the distribution of the number of venues associated with each station can be seen in Figure 6.

	Main Category	Count	Percentage
0	Food	1988	52.75
1	Shop & Service	737	19.55
2	Nightlife Spot	324	8.60
3	Outdoors & Recreation	311	8.25
4	Arts & Entertainment	231	6.13
5	Travel & Transport	153	4.06
6	Professional & Other Places	22	0.58
7	College & University	3	0.08

Table 1: Frequency of Venue Main Categories for All Venues




```
Mininum Venues Retrieved: 2
Maximum Veneus Retrieved: 100
```

Figure 6: Distribution of Number of Venues Retrieved per Station

Finally, S- and U-Bahn stations that returned less than 10 venues were dropped since including them in the cluster analysis would produce inaccurate labels, leaving 79 stations in total.

3.9 Applying the K-Means Algorithm

To prepare the data for clustering, the venue subcategories must be one-hot encoded and aggregated so that each station would be represented by a single row and each venue subcategory would become a column displaying the average frequency of occurrence. Using these frequencies, a new dataframe was also created to show the 5 most common venue subcategories for each station. The K-Means algorithm was then applied to the data using 3 clusters as the hyperparameter, and the resulting cluster labels plus the geographical information of each station were added to the aforementioned dataframe.

3.10 Mapping and Characterizing Clusters

The final clusters were then visualized using Folium and the localities they are situated in were foregrounded for emphasis, as seen in Figure 7.

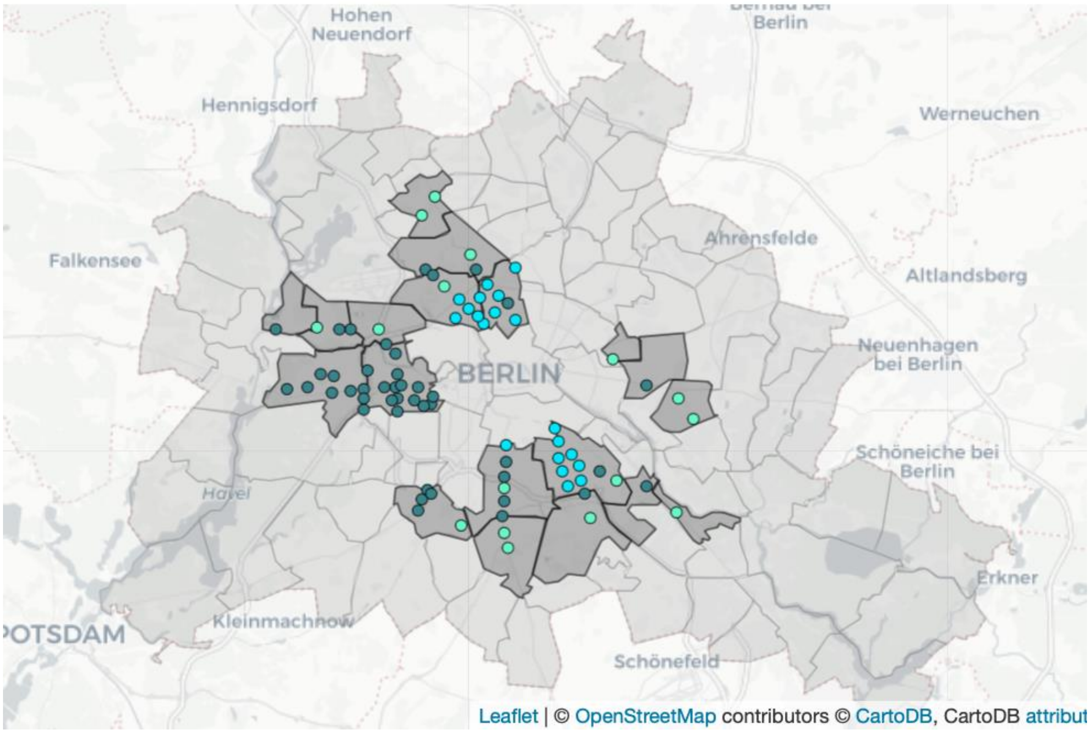


Figure 7: Final Clusters (Light Blue: Cluster 0, Dark Blue: Cluster 1, Light Green: Cluster 2)

Lastly, to facilitate the process of describing each cluster type, the cluster labels were also added to the venues dataframe created in step 3.8 to allow for grouping operations on venue categories.

4. Results

The three criteria used to narrow down the 96 localities to the 20 most suitable areas for university students were:

- Proximity to universities measured by the number of universities reachable within an 8.36-kilometer radius of the locality
- Rent prices in 2019 measured in €/m^2
- Proportion of young adults living in the locality in 2019 out of the locality's total population

These 20 localities can be seen in Table 2 below. The locality with the highest proximity score is Gesundbrunnen with a count of 27 universities, while the lowest is Niederschöneweide with a count of 8. The lowest rent prices of 8.3 €/m^2can be found in Haselhorst, while the locality with the highest rent prices is Charlottenburg at 11.1 €/m2. The locality with the youngest demographics is Wedding with a young adult proportion of 24.66%, while the locality with lowest proportion of young adults is Baumschulenweg at 15.58%.

	Locality Name	Count	Rent Prices 2019(€)	Young Adult Proportion(%)
0	Gesundbrunnen	27	10.0	24.07
1	Charlottenburg	27	11.1	18.02
2	Tempelhof	26	10.2	18.27
3	Wedding	25	11.0	24.66
4	Charlottenburg-Nord	24	8.8	16.91
5	Reinickendorf	24	9.5	19.73
6	Neukölln	23	8.9	20.23
7	Steglitz	23	10.7	16.59
8	Fennpfuhl	17	9.5	17.88
9	Britz	16	9.3	17.51
10	Westend	16	11.0	17.13
11	Borsigwalde	14	9.2	17.91
12	Haselhorst	13	8.3	16.31
13	Baumschulenweg	13	9.6	15.58
14	Lichtenberg	13	10.0	20.39
15	Siemensstadt	13	10.0	19.12
16	Mariendorf	12	9.4	16.18
17	Friedrichsfelde	11	8.4	17.00
18	Wittenau	10	9.4	16.10
19	Niederschöneweide	8	9.3	20.33

Table 2: 20 Best Localities for University Students

Within these 20 localities, there were a total of 97 S- and U-Bahn stations, which were the origin points when searching for venue data using the Foursquare API. In other words, there were 97 potential neighborhoods.

Out of these 97 potential neighborhoods, 18 were dropped from the cluster analysis because they had less than 10 venues within a 700-meters walking catchment, leaving a total of 79 stations. Due to this change, only 19 localities were represented in the final list of train stations used for clustering. The locality that was dropped from the cluster analysis was Borsigwalde.

Charlottenburg has the highest number of stations with all 17 stations belonging to Cluster 1. Within a locality, it is more common for the stations to share the same cluster type as opposed to having a variety. The clearest exception would be Tempelhof, Neukölln, and Wedding, which possess all three cluster types.

Cluster Zero

Cluster Zero has 20 stations in total (Figure 8). 8 stations are in Neukölln, 6 in Wedding, 5 in Gesundbrunnen, and 1 in Tempelhof. On average, there are 63 venues within 700 meters of each station. Most of the venues in Cluster Zero are categorized as food (53.6%), followed by nightlife spot (17.6%) and shop & service (12.8%). The 1st most common venue subcategory in Cluster Zero is cafes, the 2nd most common is bars, and the 3rd most common is pubs.

Cluster One

Cluster One has 43 stations, the highest of all the clusters (Figure 9). This cluster is extremely spaced out compared to Cluster Zero. 17 stations are in Charlottenburg, 7 in Westend, 4 in Tempelhof, 4 in Steglitz, 2 in Reinickendorf, 2 in Siemensstadt, and 1 each in Lichtenberg, Haselhorst, Neukölln, Gesundbrunnen, Baumschulenweg, Wedding, and Britz. On average, there are 49 venues within 700 meters of each station. Most of the venues in Cluster One are categorized as food (55.1%), followed by shop & service (20.6%) and outdoors & recreation (7.1%). The 1st most common venue subcategory is supermarkets, the 2nd most common is Italian restaurants, and the 3rd most common is cafes.

Cluster Two

Finally, there are 16 stations in Cluster Two (Figure 10). Like Cluster One, this cluster is also spatially distributed. 2 stations are in Wittenau, 2 in Friedrichsfelde, 2 in Mariendorf, and 1 each in Fennpfuhl, Tempelhof, Charlottenburg-Nord, Reinickendorf, Niederschöneweide, Steglitz, Siemensstadt, Neukölln, Wedding, and Britz. On average, there are 17 venues within 700 meters of each station. Most of the venues in Cluster Two are categorized as shop & service (39.6%), followed by food (35.2%) and outdoors & recreation (13.9%). The 1st most common venue subcategory is supermarkets, the 2nd most common is parks, and the 3rd most common is drugstores.

5. Discussion

Based on the characteristics of the three clusters, the following recommendations can be made to prospective university students:

- Stations in Cluster Zero serve as suitable neighborhoods for university students who thrive in a lively environment and would like to participate in the area's bustling social scenes. These neighborhoods are good for students who love to go out but prefer cheaper rent prices compared to areas with Berlin's top nightlife spots such as Prenzlauer Berg, Mitte, Kreuzberg, and Friedrichshain.
- In comparison to Cluster Zero, stations in Cluster One and Two are better options for university students who do not put as much priority on the nightlife scene. Both clusters offer similar venue categories such as supermarkets and restaurants. These neighborhoods are geared towards students who seek comfort and convenience.
- However, stations in Cluster Two would be more suitable than those in Cluster One for students who prefer to live in a quieter neighborhood and value their privacy.

5.1 Future Recommendations

While this project used only three criteria for selecting localities, there are several other factors that are just as important to university students who are looking for a new home. These include, but are not limited to, crime rates and personal safety, job opportunities, the price range and ratings of surrounding venues, and the socioeconomic characteristics of their neighbors. However, an important caveat to mention is the findings produced using data science may not correspond with student's perceptions and interests. There is most likely a correlation between high rent prices and high livability. Some students may be willing to pay more to live only a few blocks away from social venues and their friends or to live in a hip neighborhood.

Furthermore, housing was examined in this project only from the perspective of rent prices, but in Berlin, shared apartments (Wohngemeinschaft) and student accommodations (Studentenwohnheim) are pretty popular, which could allow for cheaper rent prices in areas that are typically more expensive. Another factor that future studies could take into consideration is incorporating bus stops and tram stations into the dataset, which could result in more unique cluster types and more accurately capture the experience of living in the outer localities.

Besides including more variables into the analysis, future studies could also optimize the number of clusters used for the K-means algorithm by applying the elbow method or the silhouette score. Finally, another potential data product to examine is a classification model that takes user inputs such as a specific university or venue category and assigns a cluster label, allowing the user to effectively narrow down their apartment search.

6. Conclusion

The project's main goals were to identify the most appealing neighborhoods in Berlin to university students and to describe the experiences of living in one of these neighborhoods using venue data. The end results of this project fulfilled these goals by creating a list of 20 localities and segmenting the train stations into three clusters with unique venue categories and characteristics. Stakeholders could apply a similar methodology to find the best neighborhoods for themselves or their clients, and this workflow could easily be generalized to other location problems such as finding the best site to open an organic produce store.

This project was a great opportunity for to me to demonstrate the data science skills that I've acquired from the IBM course and from my self-study and integrate them into a comprehensive project. I felt challenged when I pushed the limits of my skillset but satisfied as I started developing my independence as a data scientist. There's a lot more to learn, and I am excited for my journey ahead.

	Station Name	Locality	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	U Hermannplatz	Neukölln	Neukölln	52.486662	13.425104	0	Bar	Coffee Shop	Café	Pizza Place	Italian Restaurant
1	S+U Neukölln	Neukölln	Neukölln	52.469287	13.442349	0	Bar	Cocktail Bar	Plaza	Middle Eastern Restaurant	Ice Cream Shop
2	U Leopoldplatz	Wedding	Mitte	52.546481	13.359376	0	Café	Bar	Drugstore	Turkish Restaurant	Ice Cream Shop
3	U Osloer Straße	Gesundbrunnen	Mitte	52.556926	13.373265	0	Café	Bar	Soccer Field	Bakery	Doner Restaurant
4	U Nauener Platz	Wedding	Mitte	52.551512	13.367346	0	Bar	Café	Restaurant	Ice Cream Shop	Deli / Bodega
5	U Amrumer Straße	Wedding	Mitte	52.542191	13.349515	0	Café	Bar	Korean Restaurant	Plaza	Vietnamese Restaurant
6	U Karl-Marx-Straße	Neukölln	Neukölln	52.476418	13.439785	0	Café	Bar	Art Gallery	Coffee Shop	Plaza
7	U Rathaus Neukölln	Neukölln	Neukölln	52.481057	13.435227	0	Bar	Café	Coffee Shop	Italian Restaurant	Breakfast Spot
8	U Pankstraße	Gesundbrunnen	Mitte	52.552243	13.381818	0	Bar	Turkish Restaurant	Café	Hotel	Drugstore
9	U Voltastraße	Gesundbrunnen	Mitte	52.541137	13.393850	0	Café	Bakery	Ice Cream Shop	Gastropub	Flea Market
10	U Schönleinstraße	Neukölln	Friedrichshain-Kreuzberg	52.493167	13.422222	0	Bar	Coffee Shop	Café	Pizza Place	Pub
11	U Boddinstraße	Neukölln	Neukölln	52.479733	13.425763	0	Bar	Coffee Shop	Italian Restaurant	Pub	French Restaurant
12	U Leinestraße	Neukölln	Neukölln	52.473230	13.428496	0	Café	Bar	Pub	Bakery	Pizza Place
13	U Seestraße	Wedding	Mitte	52.550429	13.352127	0	Café	Bar	Supermarket	Drugstore	Middle Eastern Restaurant
14	U Reinickendorfer Straße	Wedding	Mitte	52.539884	13.370375	0	Bakery	Performing Arts Venue	Art Gallery	Bar	Bistro
15	U Platz der Luftbrücke	Tempelhof	Tempelhof-Schöneberg	52.484966	13.386332	0	Café	Italian Restaurant	Gay Bar	Historic Site	Bar
16	S+U Hermannstraße	Neukölln	Neukölln	52.467170	13.431684	0	Café	Coffee Shop	Pub	Park	Turkish Restaurant
17	S+U Wedding	Wedding	Mitte	52.542720	13.366041	0	Bar	Café	Turkish Restaurant	Bakery	Bistro
18	S Humboldthain	Gesundbrunnen	Mitte	52.544730	13.378707	0	Gym / Fitness Center	Art Gallery	Nightclub	Bar	Bakery
19	S Wollankstr.	Gesundbrunnen	Mitte	52.564962	13.393131	0	Café	Sushi Restaurant	Zoo Exhibit	Pub	Bed & Breakfast

Figure 8: Cluster Zero

	Station Name	Locality	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	U Kurfürstendamm	Charlottenburg	Charlottenburg-Wilmersdorf	52.503751	13.331400	1	Hotel	Zoo Exhibit	Café	Clothing Store	Italian Restaurant
1	U Schloßstraße	Steglitz	Steglitz-Zehlendorf	52.461172	13.324817	1	Café	Sushi Restaurant	Clothing Store	Supermarket	Bakery
2	U Walther-Schreiber-Platz	Steglitz	Tempelhof-Schöneberg	52.464866	13.328754	1	Supermarket	Sushi Restaurant	Café	Vietnamese Restaurant	Korean Restaurant
3	U Grenzallee	Britz	Neukölln	52.463505	13.444808	1	Bakery	Hotel	Supermarket	Rental Car Location	Hardware Store
4	U Zitadelle	Haselhorst	Spandau	52.537511	13.217607	1	Gym / Fitness Center	Shoe Store	Go Kart Track	Beer Store	Big Box Store
5	U Siemensdamm	Siemensstadt	Spandau	52.537014	13.273305	1	Supermarket	Park	Hotel	Miscellaneous Shop	Thai Restaurant
6	U Adenauerplatz	Charlottenburg	Charlottenburg-Wilmersdorf	52.500074	13.307329	1	Hotel	Italian Restaurant	Plaza	Bakery	Sushi Restaurant
7	U Rohrdamm	Siemensstadt	Spandau	52.536932	13.264164	1	Italian Restaurant	Hotel	Supermarket	Plaza	Sculpture Garden
8	U Mierendorffplatz	Charlottenburg	Charlottenburg-Wilmersdorf	52.525966	13.305697	1	Supermarket	Café	Hotel	Asian Restaurant	Italian Restaurant
9	U Richard-Wagner-Platz	Charlottenburg	Charlottenburg-Wilmersdorf	52.517142	13.307203	1	Supermarket	Café	Italian Restaurant	Hotel	Pizza Place
10	U Bismarckstraße	Charlottenburg	Charlottenburg-Wilmersdorf	52.511502	13.305267	1	Café	Italian Restaurant	Thai Restaurant	Chinese Restaurant	Pizza Place
11	U Wilmersdorfer Straße	Charlottenburg	Charlottenburg-Wilmersdorf	52.505822	13.306866	1	Italian Restaurant	Café	Thai Restaurant	Coffee Shop	Vietnamese Restaurant
12	U Magdalenenstraße	Lichtenberg	Lichtenberg	52.512202	13.489419	1	Bakery	Supermarket	Pizza Place	Italian Restaurant	Hotel
13	U Olympia-Stadion	Westend	Charlottenburg-Wilmersdorf	52.517037	13.250482	1	Park	Bakery	Italian Restaurant	Gourmet Shop	Fast Food Restaurant
14	U Neu-Westend	Westend	Charlottenburg-Wilmersdorf	52.516398	13.259892	1	Italian Restaurant	Café	Ice Cream Shop	Supermarket	Indian Restaurant
15	U Theodor-Heuss-Platz	Westend	Charlottenburg-Wilmersdorf	52.509787	13.272959	1	Hotel	Italian Restaurant	Supermarket	Indian Restaurant	Bakery
16	U Kaiserdamm	Westend	Charlottenburg-Wilmersdorf	52.509960	13.281949	1	Hotel	Supermarket	Café	Asian Restaurant	Restaurant
17	U Sophie-Charlotte-Platz	Charlottenburg	Charlottenburg-Wilmersdorf	52.510959	13.297444	1	Café	Italian Restaurant	Vietnamese Restaurant	Pizza Place	Trattoria/Osteria
18	U Deutsche Oper	Charlottenburg	Charlottenburg-Wilmersdorf	52.511816	13.309401	1	Café	Italian Restaurant	Thai Restaurant	Chinese Restaurant	Pizza Place
19	U Ernst-Reuter-Platz	Charlottenburg	Charlottenburg-Wilmersdorf	52.511490	13.322501	1	Hotel	Coffee Shop	Italian Restaurant	Indian Restaurant	Café
20	U Uhlandstraße	Charlottenburg	Charlottenburg-Wilmersdorf	52.502731	13.326215	1	Hotel	Café	Italian Restaurant	German Restaurant	Clothing Store
21	U Franz-Neumann-Platz	Reinickendorf	Reinickendorf	52.563678	13.364480	1	Supermarket	Plaza	Pizza Place	Park	Pastry Shop
22	U Kurt-Schumacher-Platz	Reinickendorf	Reinickendorf	52.563472	13.327310	1	Italian Restaurant	Bakery	Rental Car Location	Supermarket	Seafood Restaurant
23	U Afrikanische Straße	Wedding	Mitte	52.561070	13.333136	1	Supermarket	Italian Restaurant	Plaza	Bakery	Rental Car Location
24	U Ullsteinstraße	Tempelhof	Tempelhof-Schöneberg	52.453439	13.384752	1	Supermarket	Clothing Store	Drugstore	Doner Restaurant	Gym / Fitness Center
25	U Paradestraße	Tempelhof	Tempelhof-Schöneberg	52.478131	13.386705	1	Historic Site	Event Space	Bakery	Nightclub	Doner Restaurant
26	U Kaiserin-Augusta-Straße	Tempelhof	Tempelhof-Schöneberg	52.460501	13.384886	1	Supermarket	Café	Park	Doner Restaurant	Drugstore
27	S+U Tempelhof	Tempelhof	Tempelhof-Schöneberg	52.470683	13.385735	1	Supermarket	Italian Restaurant	Fried Chicken Joint	Vietnamese Restaurant	Skate Park
28	S+U Jungfernheide Bhf	Charlottenburg	Charlottenburg-Wilmersdorf	52.530264	13.299045	1	Supermarket	Playground	Lake	Mexican Restaurant	Café
29	S+U Gesundbrunnen Bhf	Gesundbrunnen	Mitte	52.549025	13.388026	1	Drugstore	Ice Cream Shop	Turkish Restaurant	Bookstore	Trail
30	S+U Zoologischer Garten Bhf	Charlottenburg	Charlottenburg-Wilmersdorf	52.506871	13.332908	1	Zoo Exhibit	Hotel	Clothing Store	German Restaurant	Cocktail Bar
31	S Olympiastadion	Westend	Charlottenburg-Wilmersdorf	52.511567	13.240752	1	Soccer Stadium	Lounge	Plaza	Restaurant	Pool
32	S Westkreuz	Charlottenburg	Charlottenburg-Wilmersdorf	52.501141	13.283018	1	Hotel	Italian Restaurant	Boarding House	Ice Cream Shop	Bakery
33	S Baumschulenweg	Baumschulenweg	Treptow-Köpenick	52.466723	13.489648	1	Supermarket	Drugstore	Café	Italian Restaurant	Bakery
34	S Sonnenallee	Neukölln	Neukölln	52.473252	13.456166	1	Supermarket	Breakfast Spot	Bar	Bakery	Café
35	S Westend	Charlottenburg	Charlottenburg-Wilmersdorf	52.518644	13.285001	1	Café	Supermarket	German Restaurant	Furniture / Home Store	Bakery
36	S Messe Nord/ ICC	Charlottenburg	Charlottenburg-Wilmersdorf	52.506443	13.283025	1	Hotel	Bakery	Doner Restaurant	Café	Asian Restaurant
37	S Savignyplatz	Charlottenburg	Charlottenburg-Wilmersdorf	52.504960	13.319801	1	Hotel	Café	Italian Restaurant	German Restaurant	Coffee Shop
38	S Pichelsberg	Westend	Charlottenburg-Wilmersdorf	52.510545	13.226730	1	Stadium	Italian Restaurant	Austrian Restaurant	Hockey Rink	Hotel
39	S Heerstr.	Westend	Charlottenburg-Wilmersdorf	52.508584	13.259373	1	Café	Supermarket	Thai Restaurant	Italian Restaurant	Chinese Restaurant
40	S Charlottenburg	Charlottenburg	Charlottenburg-Wilmersdorf	52.505141	13.303459	1	Italian Restaurant	Café	Vietnamese Restaurant	Thai Restaurant	German Restaurant
41	S Feuerbachstr.	Steglitz	Steglitz-Zehlendorf	52.463567	13.332394	1	Café	Clothing Store	Supermarket	Italian Restaurant	Vietnamese Restaurant
42	S Rathaus Steglitz	Steglitz	Steglitz-Zehlendorf	52.455950	13.322454	1	Sushi Restaurant	Pub	Supermarket	Drugstore	Trattoria/Osteria

Figure 9: Cluster One

	Station Name	Locality	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	U Blaschkoallee	Britz	Neukölln	52.452399	13.449005	2	Supermarket	Gas Station	Liquor Store	Eastern European Restaurant	Pet Store
1	U Paulsternstraße	Siemensstadt	Spandau	52.538087	13.248574	2	Supermarket	Furniture / Home Store	Asian Restaurant	Sandwich Place	Hardware Store
2	U Jakob-Kaiser-Platz	Charlottenburg-Nord	Charlottenburg-Wilmersdorf	52.536973	13.293643	2	Supermarket	Mexican Restaurant	Rock Climbing Spot	Bakery	Soccer Field
3	U Friedrichsfelde	Friedrichsfelde	Lichtenberg	52.505884	13.512771	2	Supermarket	Drugstore	Bakery	Zoo Exhibit	Playground
4	U Tierpark	Friedrichsfelde	Lichtenberg	52.497225	13.523606	2	Supermarket	Zoo Exhibit	Drugstore	Snack Place	Hotel
5	U Rathaus Reinickendorf	Wittenau	Reinickendorf	52.588206	13.325549	2	Supermarket	Trattoria/Osteria	Concert Hall	Eastern European Restaurant	Electronics Store
6	U Residenzstraße	Reinickendorf	Reinickendorf	52.570550	13.360859	2	Supermarket	Bank	Trattoria/Osteria	Market	Mini Golf
7	U Rehberge	Wedding	Mitte	52.556117	13.341980	2	Supermarket	Bar	Restaurant	Ice Cream Shop	Café
8	U Alt-Tempelhof	Tempelhof	Tempelhof-Schöneberg	52.465919	13.385777	2	Supermarket	Park	Fried Chicken Joint	Italian Restaurant	Bakery
9	U Westphalweg	Mariendorf	Tempelhof-Schöneberg	52.445790	13.385542	2	Supermarket	German Restaurant	Miscellaneous Shop	Fried Chicken Joint	Bank
10	U Alt-Mariendorf	Mariendorf	Tempelhof-Schöneberg	52.439104	13.387861	2	Supermarket	Bakery	German Restaurant	Fast Food Restaurant	Park
11	S+U Wittenau	Wittenau	Reinickendorf	52.596279	13.334943	2	Supermarket	Park	Candy Store	Italian Restaurant	Mediterranean Restaurant
12	S Schöneweide Bhf	Niederschöneweide	Treptow-Köpenick	52.455034	13.511321	2	Supermarket	Falafel Restaurant	Furniture / Home Store	Pharmacy	Gas Station
13	S Köllnische Heide	Neukölln	Neukölln	52.469157	13.467509	2	Supermarket	Trattoria/Osteria	Drugstore	Eastern European Restaurant	Playground
14	S Storkower Str.	Fennpfuhl	Lichtenberg	52.524094	13.465510	2	Supermarket	Park	Soccer Field	Pizza Place	Bakery
15	S Südende	Steglitz	Steglitz-Zehlendorf	52.449035	13.353724	2	Supermarket	Italian Restaurant	Burrito Place	Korean Restaurant	Gym

Figure 10: Cluster Two