Data Science Lab

KPMG

Project Report

Analysis of Viennese Traffic Counts

By:

Michael Bösch, Paul Ladstätter, Anna Macalik, Margaritha Wanker

4894 Data Science Lab
Summerterm 2025

Abstract

Urban traffic patterns are constantly changing, influenced by socio-economic shifts, infrastructure developments and global events such as the COVID-19 pandemic. In this project, conducted in collaboration with KPMG, the historical traffic count data collected from multiple permanent counting stations across Vienna since 2016 will be analyzed. The primary goal is to identify significant trends and patterns over time and across districts, and to build predictive models that forecast future traffic developments.

To achieve these goals this project leverages the Dauerzählstellen dataset, public transportation data, population growth data and Covid19 data.

Dauerzählstellen, permanent counting points, are specific traffic monitoring locations where vehicle counts are continuously recorded over a long period of time. These points are equipped with an automated counting technology that is embedded in or above roads. This allows for a systemic collection of traffic volume data without interruption.

The analysis not only provides insights into how urban mobility has changed in recent years but also supports evidence-based decision-making for future transportation planning and policy development.

Contents

1. Introduction

This project focuses on analyzing traffic count data from Vienna to derive insights on traffic in Vienna and to develop predictive models of the future traffic volume. The primary data source is the permanent counting point dataset, which provides monthly traffic counts at various locations in the city since 2016. To enrich the analysis, the dataset will be enhanced by integrating additional data sources such as public transportation infrastructure, population data and Covid19 data. The goal is to examine historical traffic patterns and build models that forecast future traffic developments, thereby supporting KPMG in understanding mobility trends and making informed decisions related to urban transportation.

The project follows a structured methodology comprising several key phases. It begins with Exploratory Data Analysis, which includes analyzing traffic distribution across time and districts, identifying trends, seasonality and anomalies, as well as handling missing values and data inconsistencies. In the next phase, Data Engineering and Feature Creation, geographic coordinates will be standardized into WGS84 format and multiple datasets will be merged while ensuring consistency.

Following this, various predictive modeling techniques will be employed. These include Prophet and SARIMA models, which will be developed and compared to forecast traffic volumes. Clustering techniques will also be used to identify patterns in peak traffic conditions. Model performance will be evaluated using relevant metrics and cross-validation to ensure robustness. The results of our forecasting will then be integrated into an interactive dashboard that provides a transparent interface for stakeholders to explore historical data, compare the performance of different forecast models, and analyze the final ensemble prediction for any traffic station in Vienna.

2. Coordinates

The permanent counting point dataset, provided by “Magistratsabteilung 46 – Verkehrsorganisation und technische Verkehrsangelegenheiten”, uses EPSG:31256 as its coordinate reference system. EPSG:31256 is highly suitable for local-scale geospatial analysis, especially in the contexts of infrastructure planning or traffic monitoring. However, this coordinate system is not compatible with global mapping services, such as OpenStreetMap (OSM) or Google Maps, which are based on the WGS84 reference system. To ensure interoperability between the Viennese traffic count data and global datasets like public transportation networks from OSM, the coordinates from EPSG:31256 had to be transformed into WGS84. Especially since packages like osmdata or rosm were used, that are based on the WGS84 coordinate system.

3. Traffic Trends - City Centre vs. Suburbs

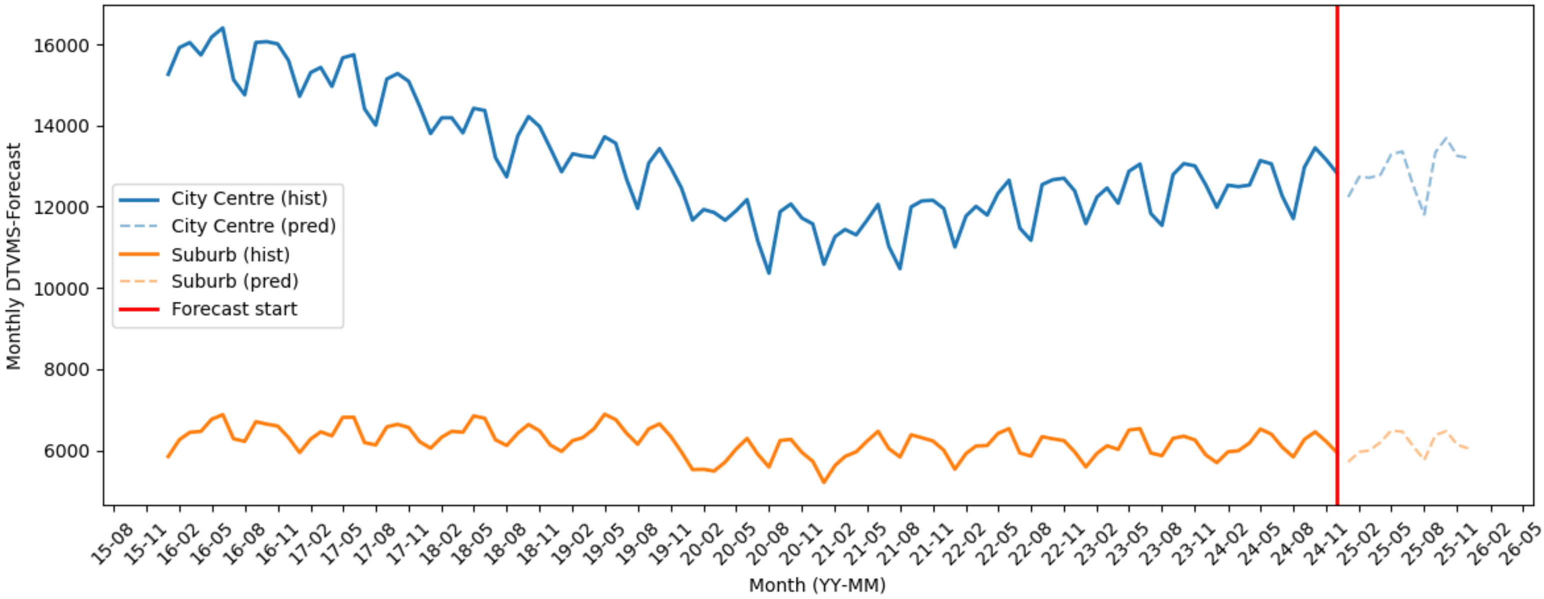

Figure 1 Traffic Volume: Historical and Prophet Forecast by Urban Zone

To project the monthly Traffic for both the City Centre and the Suburbs of Vienna and predict the upcoming traffic of 2025, first the historical data is merged with the station locations to identify the City Centre (districts 1 through 9) and the Suburbs (districts 10 through 23). The process begins with monthly averages for each weekday. These averages are used to reconstruct a complete daily time series by assigning each day the average corresponding to its weekday. This reconstructed dataset provides a consistent daily view based on limited monthly information. A time series forecasting model, Prophet, is then applied to the daily data for each Urban Zone, using data up to the last available month of February. The model is configured with weekly and yearly seasonality to capture patterns such as weekday and weekend differences and seasonal shifts, while daily seasonality is disabled because of our limited data. A 365-day forecast is generated, and the resulting daily predictions are displayed in the final monthly forecast. In the line graph in Figure 1, solid lines display the observed monthly averages, based on the historical data, up through February 2025. Dashed, semi-transparent lines then show the Prophet forecasts from March 2025 through February 2026 with the red vertical line marking the transition from historical data to model predictions.

As observed in Figure 1, the City Centre curve falls from approximately 1.600.000 DTVMS in mid 2016 to a low of about 1.050.000 in early 2021, then steadily climbs to around 1.300.000 by February 2025. The 2025–26 forecast extends the recovery from the covid collapse, fluctuating roughly between 1.200.000 and 1.400.000 and preserving the summer and winter dips that Prophet learned. The Suburb curve remains relatively flat, between about 550.000 and 680.000 DTVMS historically and its forecast carries forward the same gentle seasonal pattern, fluctuating around 560.000 to 670.000.

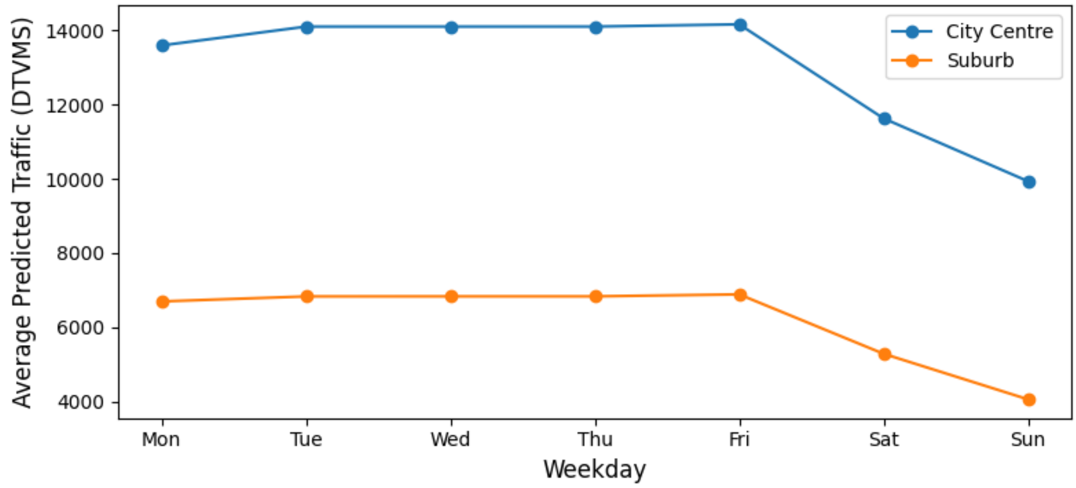

Figure 2 Average Monthly Traffic Volumes by Urban Zone

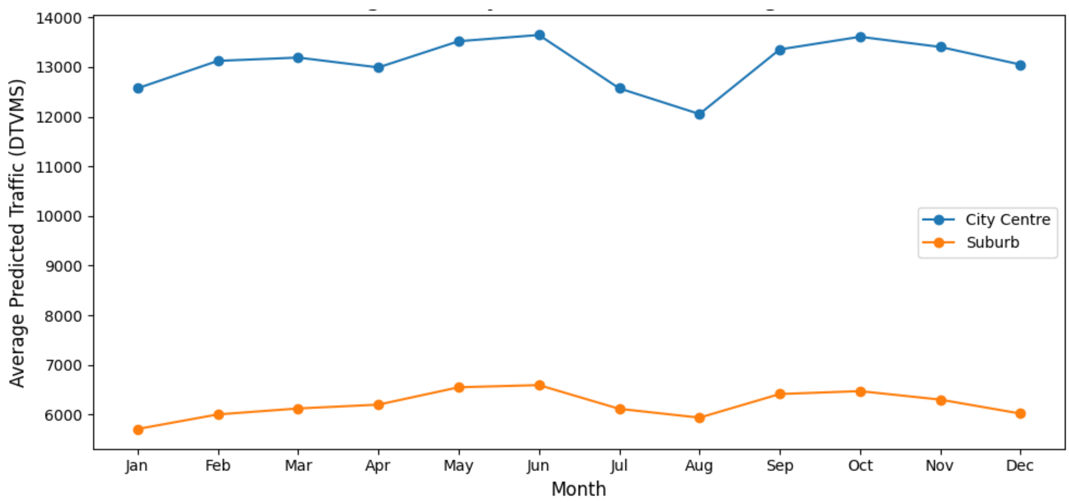

Figure 3 Average Daily Traffic by Urban Zone

Average weekly and monthly traffic patterns are derived directly from the Prophet model's daily forecasts by simple aggregation, first grouping all forecasted values by weekday (0 = Monday … 6 = Sunday) and averaging to reveal a "typical" week, then grouping by calendar month (1–12) to plot a "typical" year in months.

In the weekly pattern chart, as displayed in Figure 2, the forecasting values of the City Center as well as the Suburbs form a flat plateau from Tuesday to Thursday which is caused because the Column DTVDD combines the measurements for the three weekdays into one average variable. When looking at the graph, a sharp drop in Traffic is noticeable on weekends with the weekly minimum being averagely measured on Sundays.

The monthly-pattern chart, as seen in Figure 3, shows a clear annual cycle. City Centre traffic averages rise from 1.260.000 in January to 1.300.000 in February – March, dip slightly in April (about 1.300.000), peak around May – June (1.350.000), fall to a mid-summer low in August (1.210.000), then climb again to a smaller autumn peak in September to October before tapering off toward December and January with a steady climb in Traffic Volume in the beginning of the new year. The Suburb pattern mirrors this cycle at about half the magnitude, ranging from 570.000 in January up to 660.000 in June, dipping in August and peaking near October. Together, these plots visually confirm that the model has learned and will reproduce a stable weekday/weekend contrast and a pronounced seasonality for both central and suburban Vienna traffic.

4. Covid19 Analysis

This section presents an analysis of how the COVID-19 pandemic affected traffic volumes, focusing on two primary indicators: TVMAX, maximum traffic volume, and DTV, mean traffic volume. Using historical data, the analysis explores how traffic patterns changed during the pandemic period, including the effects of lockdowns, restrictions, and gradual reopenings.

In addition to analyzing past trends, this section includes a forward-looking simulation that models the potential impact of a similar pandemic occurring in 2030. By applying patterns observed during COVID-19 to future projections, the forecast provides insights into how traffic volumes—both TVMAX and DTV—might respond under comparable conditions.

4.1. Historic analysis

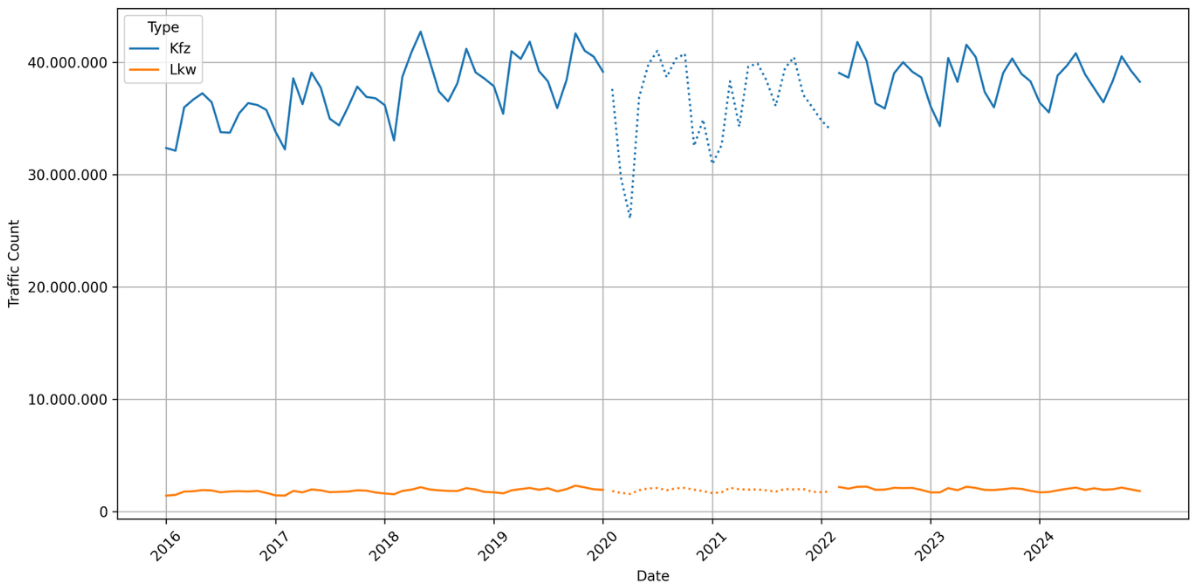

Figure 4: Time series analysis from 2016 to 2024 including a Covid19 analysis – dotted line – for vehicle types Kfz - passenger cars -and Lkw – trucks- using the maximum traffic volume (TVMAX)

Figure 4 displays the development of traffic volume over time, differentiated by vehicle type, Kfz, known as passenger cars, and Lkw, known as trucks, from 2016 to early 2025. The period affected by the COVID-19 pandemic is highlighted using a dotted line, allowing for a clear visual comparison between pre-, during- and post-pandemic traffic patterns.

In the pre-pandemic period, 2016 to late 2019, the traffic volume of passenger cars shows a regular seasonal fluctuation pattern, consistently ranging between approximately 30 and 40 million. These recurring peaks and lows suggest typical annual variation, possibly linked to holidays or weather changes. In contrast, truck traffic remains relatively constant during the same timeframe, with values generally between 100,000 and 150,000. This indicates a stable demand for transport, which is less influenced by seasonal trends.

With the start of the COVID-19 pandemic in early 2020, a sharp and immediate drop in passenger car traffic is clearly visible. The data becomes notably more irregular and volatile during the dotted line period, reflecting the dynamic nature of lockdown measures, restrictions and shifts in mobility behavior. People likely reduced commuting and personal travel drastically, especially during stricter phases of the pandemic. Meanwhile, truck traffic experienced only a slight and temporary decline, maintaining overall stability. This underlines the critical role of freight transport, which remained largely unaffected due to its essential nature in supplying goods, including during lockdowns.

In the post-pandemic phase from 2022 onwards, car traffic shows a gradual return to pre-pandemic levels, stabilizing once again around the 30 to 40 million mark. However, the seasonal patterns observed earlier appear slightly less pronounced, meaning the traffic behavior is less volatile, and potentially indicating subtle changes in travel behavior, such as an increased use of remote work or greater demand for public transportation. Truck traffic, on the other hand, continues to exhibit a steady and consistent trend, reinforcing its resilience throughout the entire observed period.

Overall, figure 4 illustrates how the pandemic significantly disrupted passenger car traffic in the short term, while freight transport remained robust. As Vienna's traffic system recovers, these patterns provide useful insights for future urban mobility planning and infrastructure resilience.

4.2. Pandemic forecast simulation

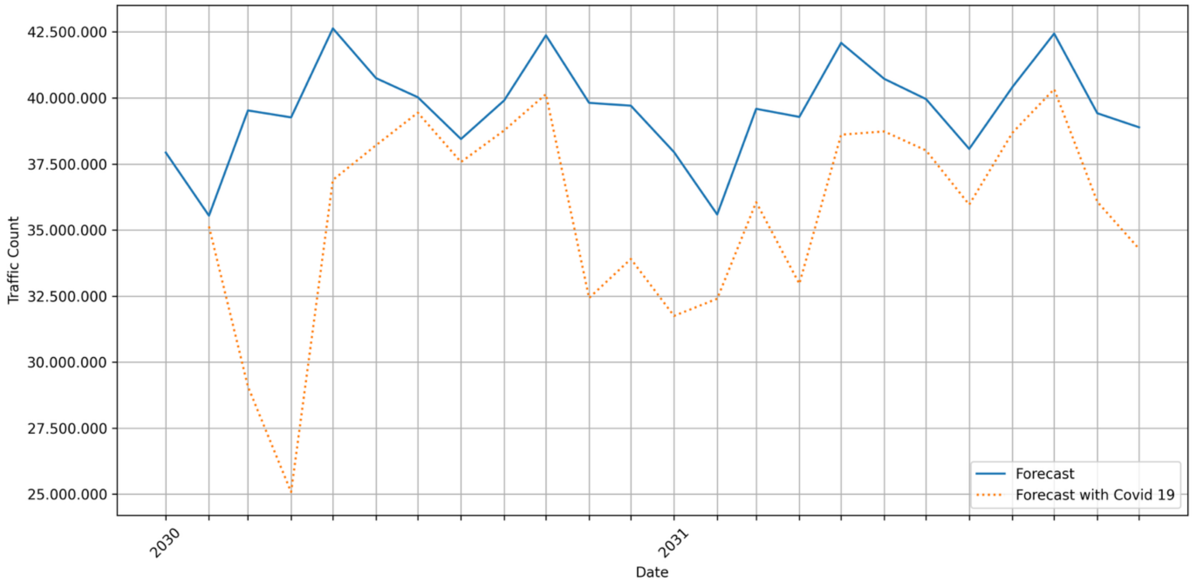

Figure 5: Covid19 simulation forecast during the years 2030 and 2032 regarding Kfz vehicles using the mean traffic volume (DTV)

Figure 5 shows the traffic forecast between 2030 and 2032 regarding Kfz vehicles. The blue line shown in the graph illustrates the traffic volume during this time assuming no pandemic occurs. In contrast, the orange dotted line visualizes the potential effects of a Covid19 pandemic if it would happen during the forecasted timeframe. The traffic volume in Vienna without Covid would remain relatively stable with values ranging from 36 to 42.5 million. These fluctuations can be explained by seasonal changes. In comparison, the scenario involving a Covid19 pandemic shows a significant decline in traffic volume. A sharp drop is evident in March 2030, coinciding with the simulation of a strict lockdown, where traffic volume falls to approximately 25 million. If the lockdown regulations ease the traffic volume recovers and rises to approximately the same level as the traffic volume expected in the no pandemic scenario. Over the following years, traffic continues to fluctuate in response to changing public health measurements. The sharper drops in the traffic volumes are subject to stricter regulation periods, such as hard lockdowns. Meanwhile the increase in traffic volume following such lockdowns suggests the lifting of restrictions.

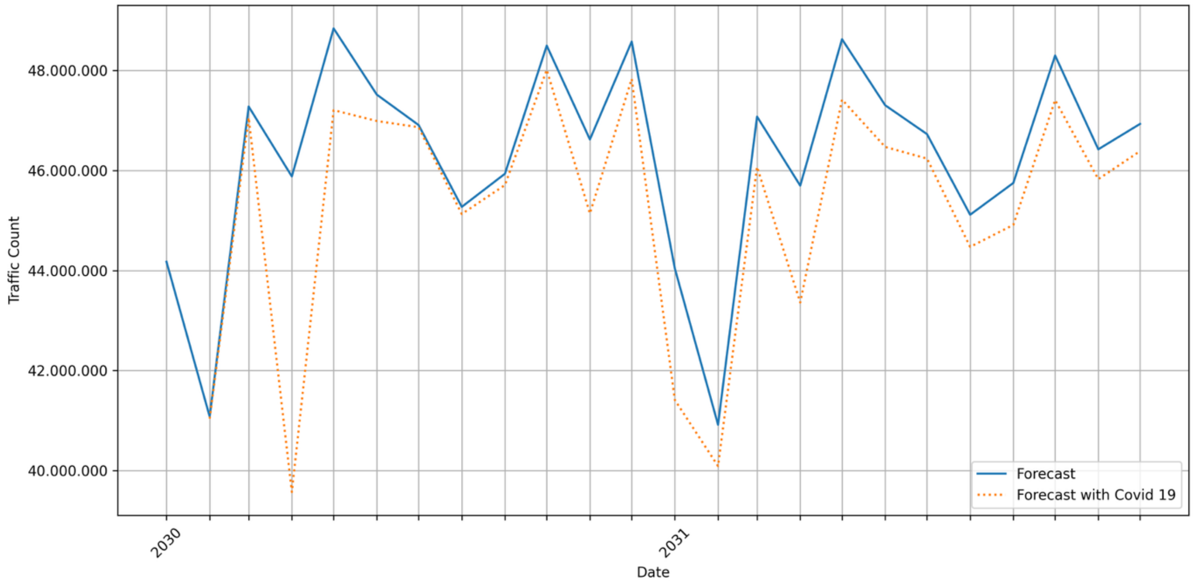

Figure 6: Covid19 simulation forecast during the years 2030 and 2032 regarding Kfz vehicles – passenger cars - using TVMAX

Figure 6 shows two traffic scenarios in Vienna from 2030 to 2032, using the value TVMAX – maximum traffic volume - as a measure of peak monthly traffic volume. In the beginning of 2030, both the normal and the pandemic-affected forecasts align closely. However, a sharp decline in the Covid19 like scenario emerges by mid 2030, indicating the onset of restrictions and reduced mobility. While traffic in the normal forecast continues to grow steadily, the simulated pandemic scenario drops significantly, reaching its lowest point during this period. As 2030 progresses, traffic in the pandemic scenario begins to recover slightly but stays consistently below the normal trend. This gap reflects ongoing behavioral changes or partial restrictions. In early 2031, a second noticeable dip appears, simulating a resurgence of the virus and therefore another strong lockdown. TVMAX again declines, reinforcing the sensitivity of traffic levels to health-related disruptions. Toward the end of 2031, both forecasts begin to stabilize, though the pandemic-influenced traffic remains lower, suggesting a shift in long-term mobility patterns. Overall, TVMAX proves to be a responsive indicator of urban mobility, capturing the impact of simulated public health crises and the slower recovery in traffic behavior over time.

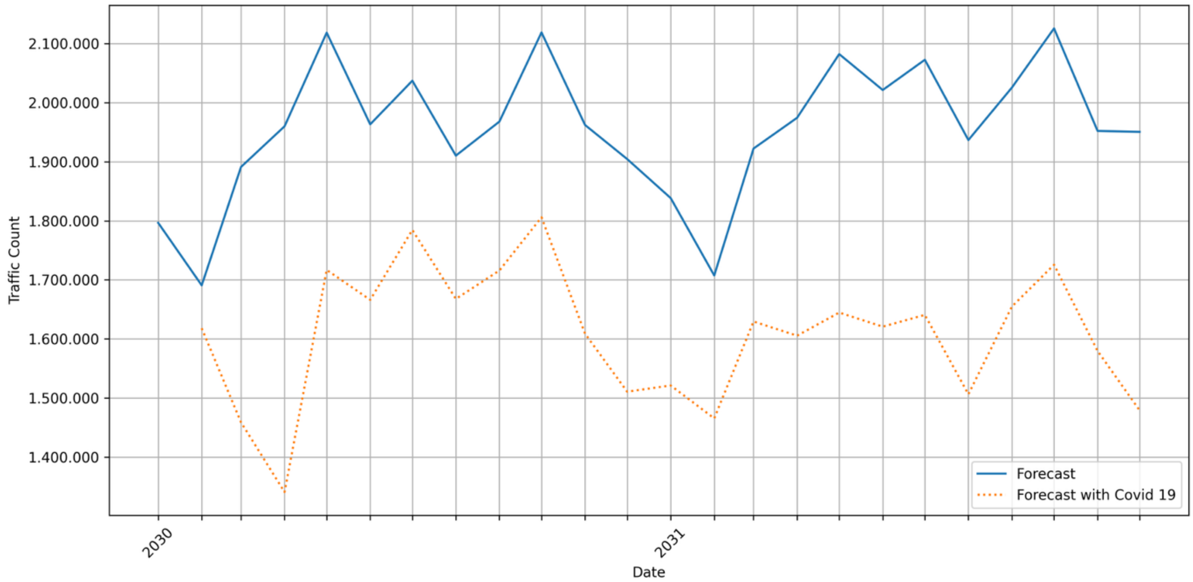

Figure 7: Covid19 simulation forecast during the years 2030 and 2032 regarding Lkw vehicles – trucks - using DTV

Figure 7 illustrates the expected monthly truck traffic volumes in Vienna from 2030 to 2032 under two different scenarios. The blue line represents the baseline forecast without any pandemic effects, while the orange dotted line shows the expected traffic forecasts with a pandemic occurring during this period.

In the scenario without a pandemic, the truck traffic volume remains relatively stable, fluctuating between approximately 1.8 and 2.1 million vehicles. These changes likely reflect seasonal variations and regular economic patterns. In contrast, the orange line analyzing the pandemic simulation shows a significant deviation. Around March 2030 the line drops sharply to about 1.35 million, indicating the simulation of a strict lockdown. This drop is followed by an increase although the traffic levels remain consistently below the non pandemic forecast. In the following months the orange graph shows lower values with ups and downs indicating that the effects of the restrictions reduce the demand and challenge the logistic operations.

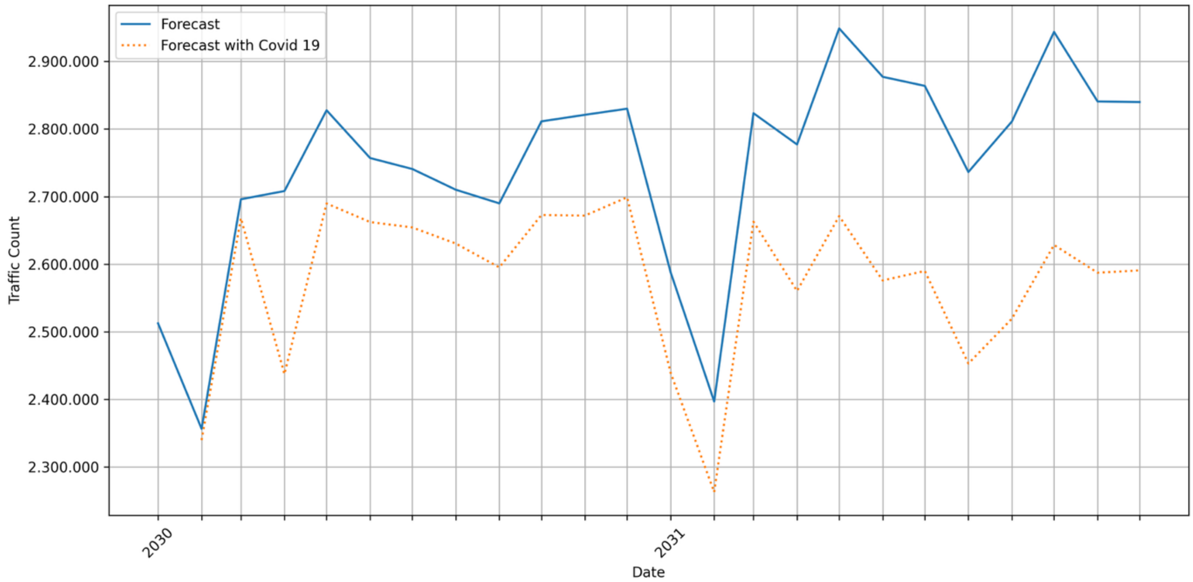

Figure 8: Covid19 simulation forecast during the years 2030 and 2032 regarding Lkw vehicles using TVMAX

Figure 8 shows forecasted traffic volumes for trucks in Vienna between 2030 and 2032, comparing a normal scenario and one influenced by a simulated Covid19 like pandemic. The value tracked represents monthly truck traffic counts. At the start of 2030, both scenarios are closely aligned, but by the second month, the Covid simulation shows a clear decline. This drop likely reflects supply chain disruptions, reduced industrial activity, and restrictions on non-essential transport. While the standard forecast predicts steady growth in truck traffic, the pandemic scenario remains consistently lower. A brief recovery occurs mid 2030, though the simulated scenario stays suppressed. A sharp dip in early 2031 suggests a second wave or renewed restrictions. This pattern mirrors disruptions in logistics and distribution experienced during real-world pandemic peaks. Toward the end of the period, traffic volumes in both scenarios stabilize, but truck volumes in the pandemic scenario remain significantly lower, indicating lasting impacts on freight movement and delivery patterns.

Overall, this forecast highlights how sensitive Lkw traffic is to economic and regulatory changes, with TVMAX offering valuable insight into pandemic-related shocks in freight activity.

5. Traffic Heatmap Analysis

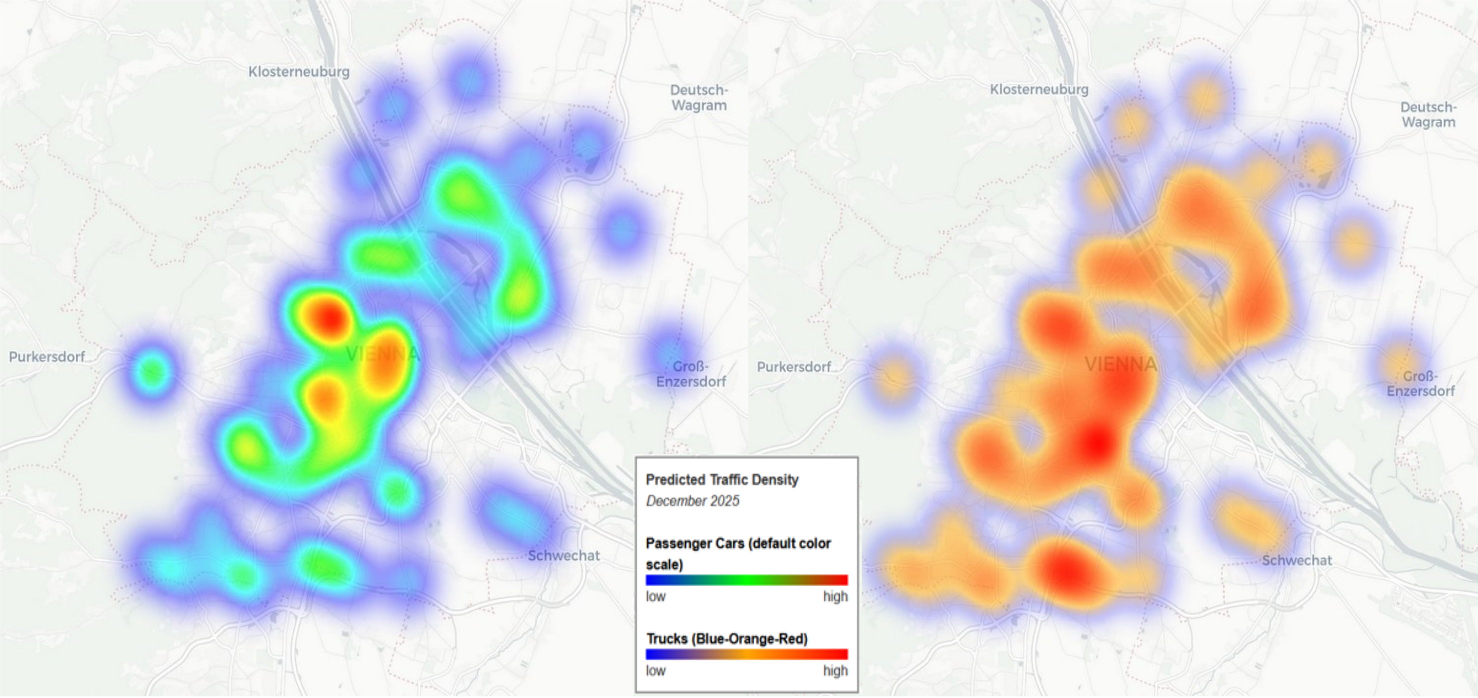

Figure 9 Predicted Traffic Density Heatmap for Vienna for Passenger Cars and Trucks (December 2025)

The predicted traffic density heatmaps for Vienna in December 2025 in figure 9 provide valuable insights into the spatial distribution of traffic for both passenger cars and trucks. These maps are based on historical vehicle count data from permanent counting stations and offer a projection of future traffic conditions using Python-based geospatial tools.

On the left side of figure 9, the heatmap for passenger cars reveals a strong concentration of traffic in the central areas of Vienna. The intensity of the colors, ranging from blue (low) to red (high), shows that the city centre is expected to experience the highest levels of passenger car traffic. This is typical for urban environments where commuters, local residents and commercial activity converge. As one moves away from the city centre, the traffic density decreases significantly, although some localized pockets of heavier traffic can still be observed along major entry and exit roads as well as near important intersections on the outskirts.

In contrast, on the right side of figure 9, the truck traffic heatmap displays a somewhat different pattern. While there is still considerable activity in the central parts of Vienna, truck traffic is more widely distributed across the city's outer areas. The colour scale for trucks, which transitions from blue to orange to red, highlights areas such as Schwechat and Groß-Enzersdorf as zones of relatively high density. These locations are associated with logistics hubs, highways and industrial zones, indicating that truck traffic is more aligned with infrastructural and logistical functions than with everyday commuting patterns. Notably, truck traffic in the very core of the city is less dominant than that of passenger cars, suggesting potential regulatory constraints or route preferences favouring outer corridors.

The city centre again emerges as the area with the most intense traffic density, but the overlay of truck and passenger car data reveals broader zones of congestion. This map shows where the two traffic types reinforce each other, leading to high overall vehicle volume. The hotspots in the central and southern districts are particularly prominent, suggesting that both cars and trucks contribute significantly to congestion in those areas. In addition, the spread of density into peripheral zones indicates that traffic issues are not limited to central Vienna but also extend toward important radial routes and suburban intersections.

While both vehicle types show high traffic near city access points, the key differences lie in the distribution and intensity of traffic across space. Passenger car traffic is more tightly clustered in the city centre and along primary commuter corridors. Truck traffic, however, is more evenly dispersed and extends into industrial and suburban areas, often following logistical infrastructure rather than population density. Despite these differences, both maps reveal overlapping congestion points, such as major ring roads and transport nodes, indicating that urban planners must consider both flows simultaneously when designing traffic management strategies.

Overall, these heatmaps illustrate that Vienna's traffic challenges in December 2025 will stem from a combination of dense inner-city car use and high truck activity on the urban periphery. The combined visualization offers a crucial planning tool, allowing decision-makers to pinpoint where interventions might most effectively reduce congestion and improve mobility for all vehicle types.

6. Generating Car Traffic Forecasts for Dashboard

In order to generate the car traffic forecasts that power our visualization dashboard, we began by assembling a complete monthly panel of average daily vehicle counts at each permanent counting station, covering January 2016 through December 2024. To explain variability in these counts, we gathered four district-level time series—commuter departure share ‘AUSPENDLER’, vehicles per 1 000 inhabitants ‘PKW_DENSITY’, district population ‘POP’ and citywide modal-split percentages. Besides the district population data, which included official forecasts, the exogenous series were extended until December 2030 using annual ARIMA(1,1,1) projections. All series were then resampled to a uniform month-start index, merged with the average daily vehicle count panel.

With this enriched dataset in hand, we fitted three separate models at each station: a seasonal time-series model with all four exogenous regressors SARIMAX, the same seasonal time-series model without regressors SARIMA, and Facebook’s additive time-series model Prophet including the same exogenous regressors. Each model trained on data up to December 2024 and produced forecasts for January 2025 through December 2030. The COVID19 pandemic was accounted for by using a dummy variable where 1 represented a pandemic phase and 0 a non-pandemic phase which the models could then learn from and adjust their forecasts accordingly.

Model	Mean RMSE	Ensemble Weight
SARIMA	1578	68.5%
Prophet	3641	29.7%
SARIMAX	62003	1.75%

Figure 10 Forecasting model performance measured in RMSE and their respective weights in the ensemble forecast.

To create the most accurate and reliable final forecast, an evaluation process was undertaken. We first assessed the performance of each model through a backtesting procedure on a two-year hold-out period, from January 2023 to December 2024. For this test, the models were trained on data only up to December 2022. They were then used to generate one-month-ahead predictions iteratively, where the model's data expands with each new month, simulating a real-world forecasting scenario.

The accuracy of these predictions was measured using the Root Mean Squared Error 'RMSE' for each model at every station. To ensure that the overall performance metric was not skewed by a few highly erratic or unpredictable counting stations, for example those affected by prolonged, unobserved events like construction and we applied a standard statistical technique. We calculated the 95th percentile of the error distribution for each model and excluded the top 5% of stations whose RMSE values exceeded this cutoff. This resulted in a more robust "trimmed-mean" RMSE that better reflects the typical performance of each model across the city.

The results of this evaluation, as summarized in Figure 10, highlight the superior performance of the SARIMA model. From these trimmed-mean RMSE values, we derived inverse-RMSE weights to form a single, weighted ensemble forecast. This process assigned the highest weight to the most accurate model, resulting in an ensemble composed of approximately 68.5% SARIMA, 29.7% Prophet, and 1.8% SARIMAX. This final ensemble projection is what powers the primary forecast presented in the dashboard, offering a balanced and resilient prediction of future traffic trends.

The final output is a unified time series for every station, spanning from 2016 to 2030. This series integrates the historical traffic counts, the individual trajectories of all three models, the final weighted ensemble forecast, and the four exogenous driver variables. This comprehensive dataset is the engine behind our dashboard, allowing users to seamlessly explore observed trends, compare alternative model paths, and view the consensus forecast within a single transparent, interactive interface.

7. Visualization Dashboard

To effectively communicate the project's findings, an interactive web-based dashboard was developed. This tool serves as the primary visual interface for exploring historical traffic data and the model-generated forecasts. The dashboard was built entirely in Python, utilizing the Dash framework by Plotly for its structure and interactivity. Geospatial mapping capabilities are provided by Dash Leaflet, with all charts and graphs rendered using Plotly Express and Plotly Graph Objects.

The dashboard is organized into two main views: a high-level map overview and a detailed analysis page for each traffic counting station.

The main page presents a map of Vienna with each traffic counting station represented by a circular marker as can be seen in figure 11. A key feature is the interactive time slider, which spans the entire dataset from January 2016 to December 2030, allowing users to navigate through the months. The size of each circle on the map corresponds to the traffic volume for the selected month, providing an immediate visual cue to traffic intensity. When a month after 2025 is selected, the date display turns red to clearly indicate that the data is model-generated.

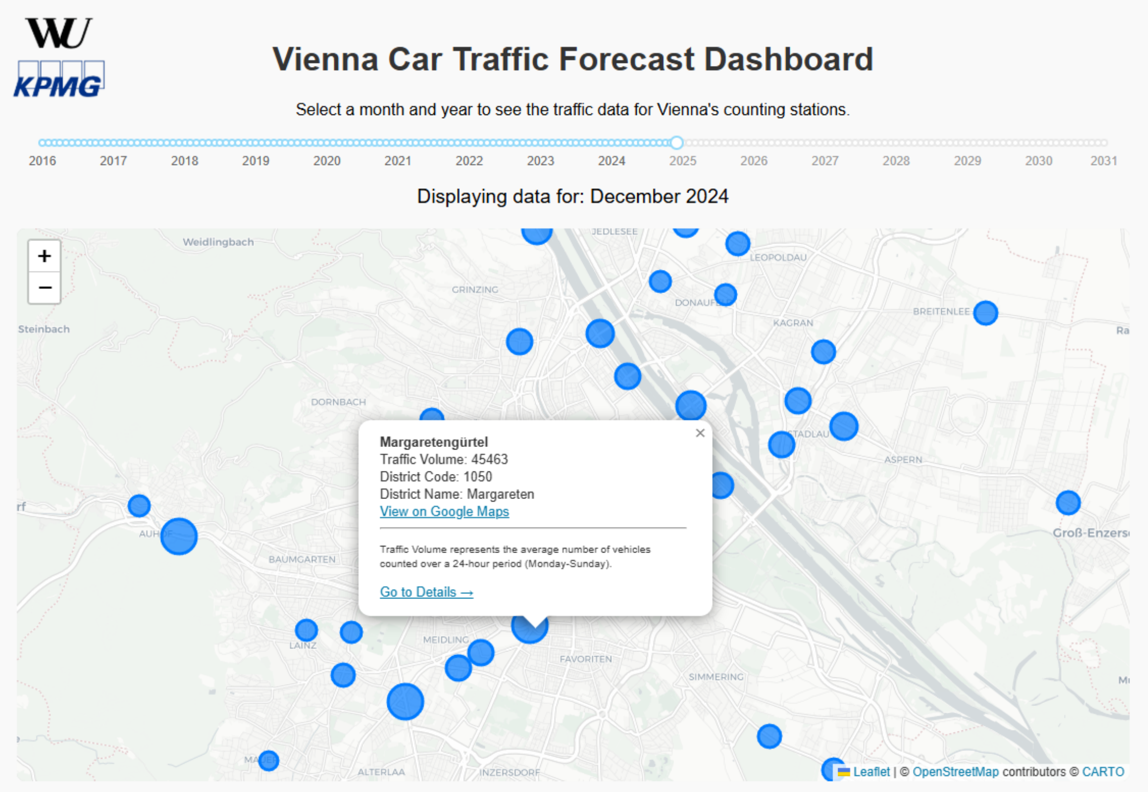

Figure 11 The main dashboard view, showing traffic counting stations across Vienna with an interactive time slider.

Clicking on a marker reveals a pop-up with key information, including the station name, traffic volume, district details, and a link to its location on Google Maps. This provides a quick summary before diving into a deeper analysis.

By clicking the "Go to Details" link in the pop-up, the user is navigated to a dedicated page for that specific station. This detail view in figure 12 features a comprehensive time series chart displaying the station's traffic volume. The solid blue line represents the actual historical data up to the end of 2024. From 2025 onwards, the chart displays multiple forecast lines. The primary prediction is the Forecast (Ensemble), shown as a solid red line, which is a weighted average of three different models. The individual forecasts from Prophet, SARIMA, and SARIMAX are also plotted as dotted lines to allow for a comparison between the different model outputs. The period of the COVID-19 pandemic is highlighted with a transparent yellow block to show its impact on traffic patterns. Below the chart, disclaimers provide context on the data and the composition of the ensemble model.

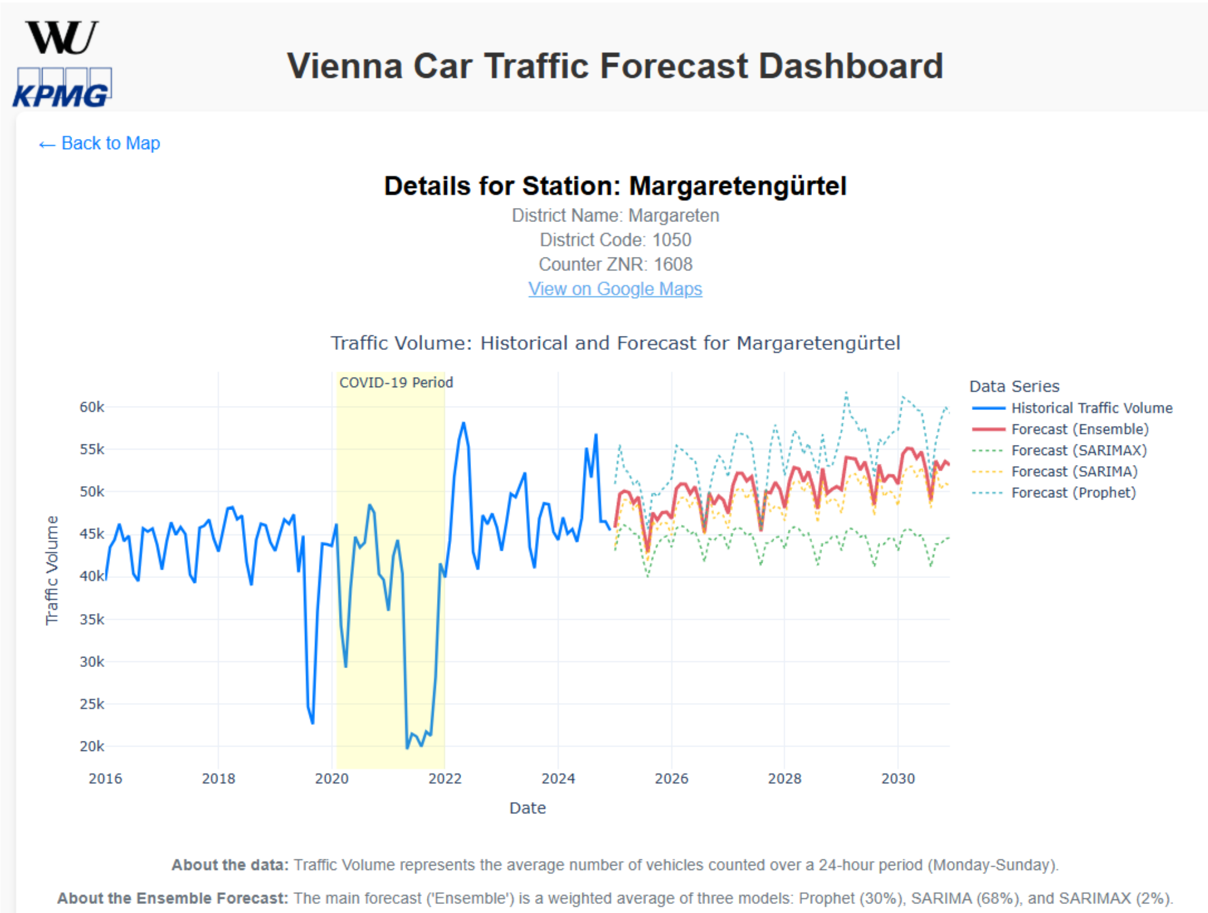

Figure 12 Detailed view for a specific station, showing historical traffic volume and multiple model forecasts.

Below the time series chart, so further down on the detail page, a second set of plots visualizes the exogenous variables used in the forecasting models in Figure 13. These charts provide deeper insight into the factors influencing traffic trends at the station's district. The time series plots for "District Population," "District Commuter Rate (%)", and "District Car Density" show their development over the entire period. The fourth chart is a stacked bar graph illustrating the yearly change in traffic share for the whole of Vienna between cars, public transport, walking, and biking. These visualizations help connect the traffic forecasts to underlying socio-economic and mobility trends.

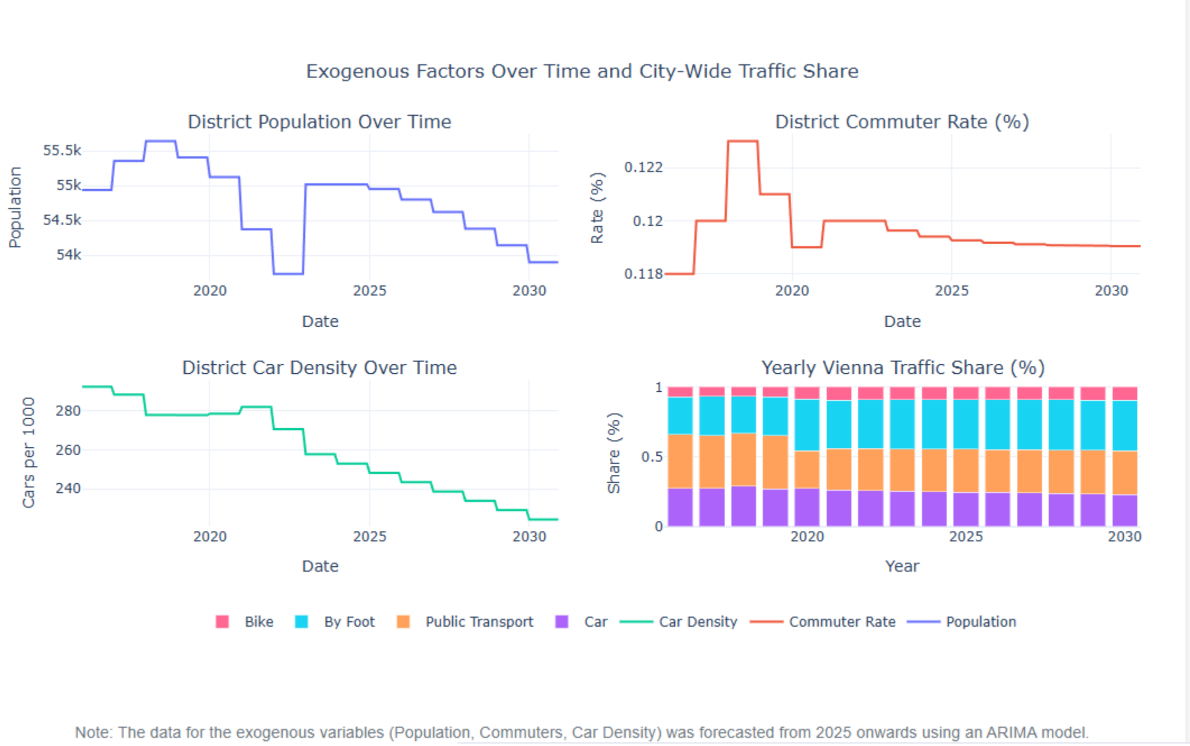

Figure 13 Exogenous variable plots for a selected station, including time series data and yearly traffic share.

This interactive dashboard successfully translates complex time series data and model outputs into an accessible and insightful tool, enabling stakeholders to explore and understand Vienna's traffic dynamics and future trends.

8. Clustering Analysis

To examine whether permanent counting points behave similarly, a clustering analysis has been conducted. The analysis provides insight into the traffic patterns of Vienna regarding Kfz vehicles – passenger cars - and Lkw vehicles - trucks.

For the clustering analysis only the combined entries from the variable "RINAME" have been considered. Therefore, no distinction between the directions have been made and the clusters are based on the two different directions a counting point measures. Furthermore, k-means was used as the clustering algorithm. K-means partitions the dataset into similar groups based on their distance between their centroids. To find the right amount of k for each analysis, the Elbow Method was used. The elbow method is a technique used to determine the optimal number of clusters in K-means clustering by plotting the within-cluster sum of squares against different values of k. The "elbow point" on the graph—where the rate of decrease in within-cluster sum of squares sharply slows—indicates the ideal number of clusters, balancing model complexity and fit.

The following six graphs plot the location of each permanent counting point based on their coordinates. Three different analyses were conducted examining different timeframes. Each of these analyses were subset and differentiated based on the vehicle type.

8.1. Clustering Analysis

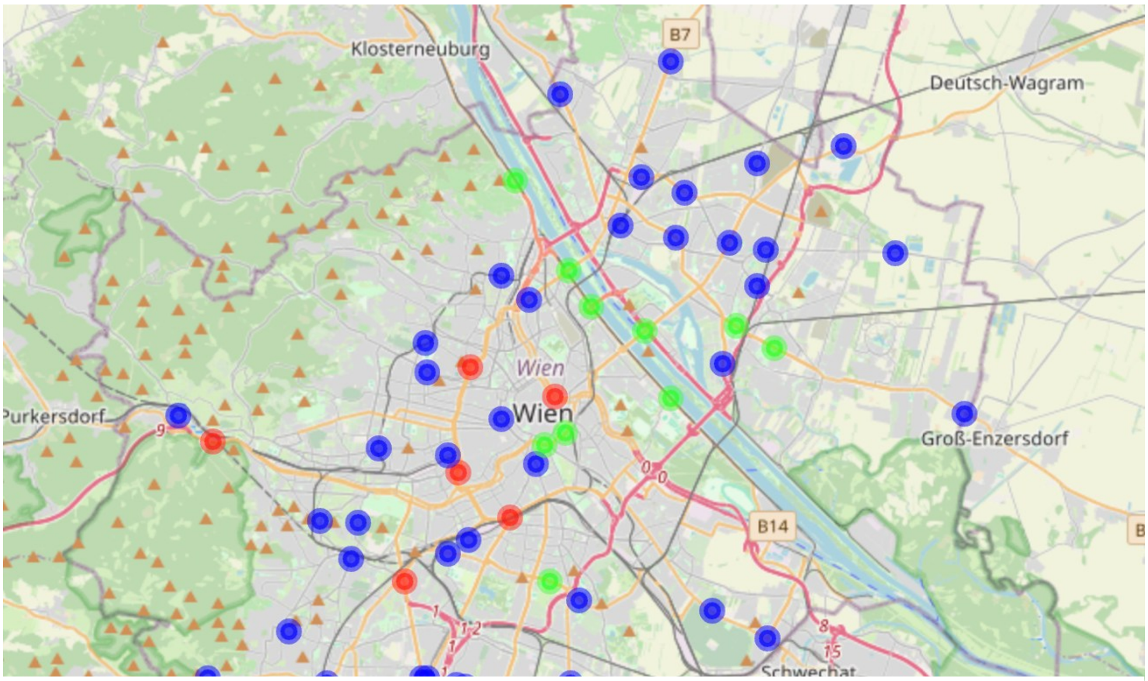

Figure 14 Cluster Analysis Kfz regarding the entire week – Cluster 1: red – highest traffic volume, Cluster 2: green – medium traffic volume, Cluster 3: blue – low traffic volume

The first analysis contains every weekday as an input. Figure 14 illustrates traffic patterns for motor vehicles and reveals a rather consistent trend in Vienna's suburban areas. For instance, the counting point "Hauptstraße" in the 14th district displays a similar traffic pattern to "Simmeringer Hauptstraße" in the 11th district. Both points are highlighted in blue, indicating that the traffic behavior in Vienna's outskirts tends to be relatively consistent and at a lower level than the rest of Vienna. This suggests that suburban areas across the city exhibit comparable traffic dynamics. However, there are notable exceptions where traffic volumes in the outskirts significantly surpass this baseline, resulting in a different clustering pattern. They are represented by the color red. The best example is the permanent counting point "Wientalstraße" in the 14th district. Despite its suburban location, it aligns more closely with the inner-city counting point "Karlsplatz" in the 1st district and therefore documents higher traffic volumes. The reason for this anomaly probably lies in the proximity of "Wientalstraße" to the major highway access point A1, which leads to substantially higher traffic volumes in that area compared to typical suburban roads.

Located between the clusters representing extreme traffic volumes is cluster 2. The cluster contains permanent traffic counting points such as "Gersthoferstraße" in the 18th district, "Hernalser Hauptsraße" in the 17th district or "Währinger Gürtel Nord/Süd". Although the 18th district is typically considered as a suburban district, it experiences a higher traffic volume pattern compared to the other outer districts. This pattern could be the result of the lack of high-capacity public transportation such as subways. Since most regions in the 18th district can only be reached by tram or car it is reasonable to assume that this absence of infrastructure leads to higher traffic volume in comparison to the better-connected 2nd district.

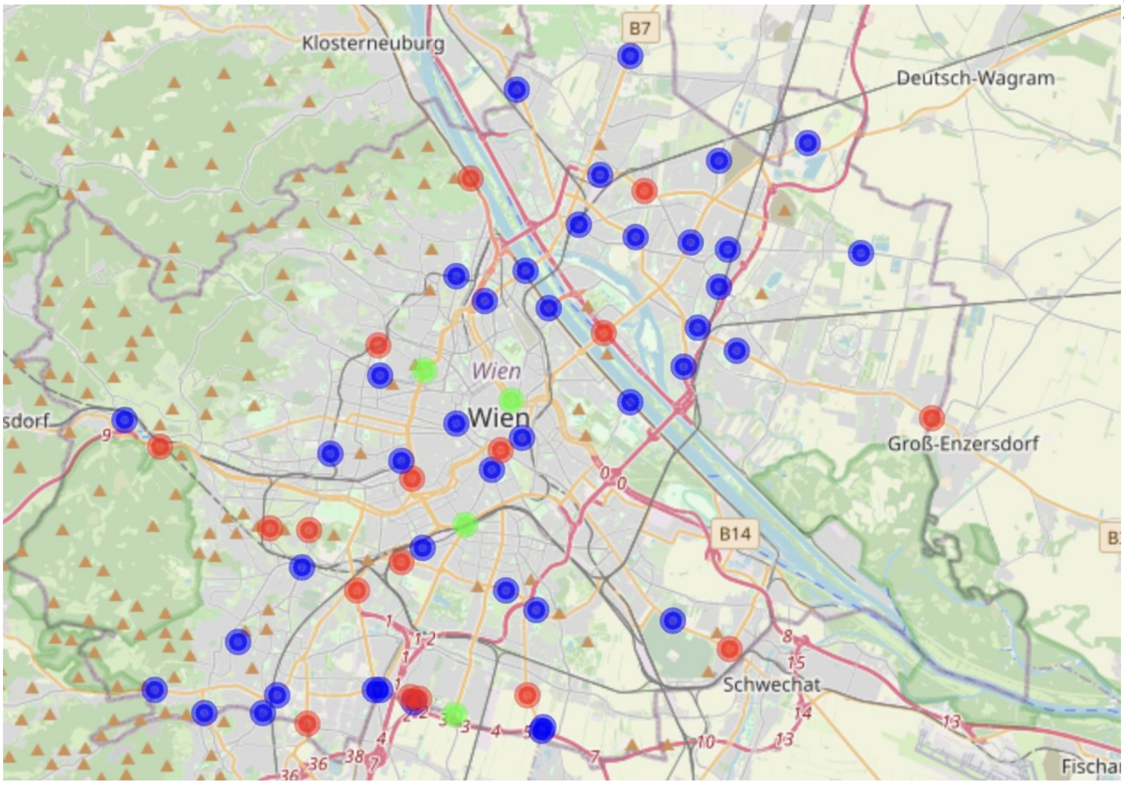

Figure 15 Cluster Analysis LKW regarding the entire week – Cluster 1: red, Cluster 2: green, Cluster 3: blue

Figure 15, the cluster analysis of heavy vehicle traffic (Lkw) of counting stations across Vienna reveals patterns in truck movement within the city. In contrast to the Kfz vehicles, not a clear distribution of traffic volume can be seen.

The majority of the counting stations are considered as cluster 3, shown in blue. These stations are broadly distributed throughout Vienna but tend to concentrate more densely in the northern and eastern parts of the city. This cluster likely represents areas with typical or average levels of heavy vehicle traffic. These may be roads that are part of well-connected arterial routes, industrial corridors, or zones that regularly support freight traffic without reaching unusually high or low volumes. As seen in the analysis above these permanent counting points are rather located in the suburbs of Vienna, such as the 13th, 18th and 23rd district.

In contrast, cluster 1 contains fewer counting stations that are more spread out, particularly toward the western and southwestern areas. This suggests that these locations experience different traffic dynamics, possibly due to higher truck volumes or less residential zoning with less limited access for heavy vehicles. These areas might also be more relevant for logistics traffic, resulting in their separation from the dominant pattern. The permanent counting point

“Wientalstraße” in the 14th district is probably one of the most important counting points for truck traffic due to its relationship with the nearby A1.

The second cluster represents a small group. These stations are mostly located closer to the city center. Their clustering indicates that they experience elevated traffic volumes—higher than average, yet not extreme enough to stand out as outliers. These locations may serve as key transitional routes connecting outer districts with the urban core or as access points to major market and delivery zones.

8.2. Clustering Analysis – Monday to Friday

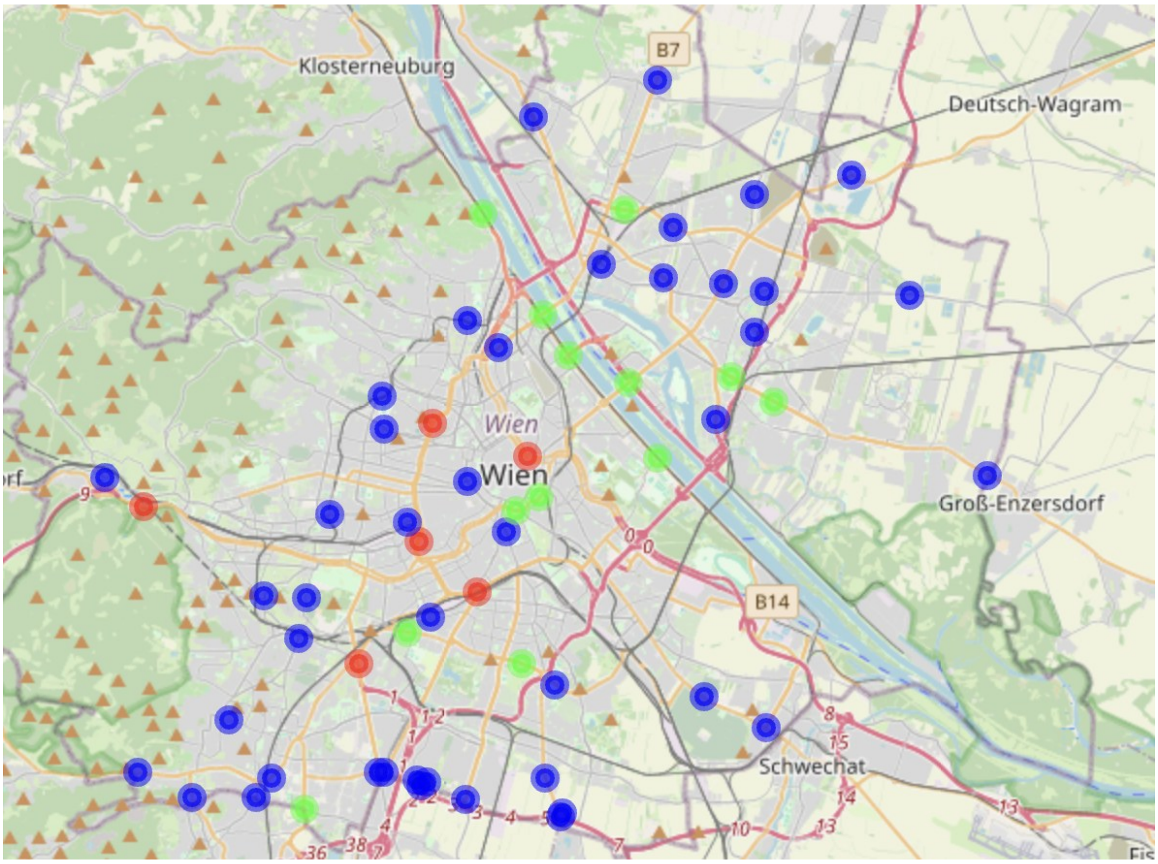

Figure 16 Cluster Analysis Kfz between Monday and Friday – Cluster 1: red, Cluster 2: green, Cluster 3: blue

The two graphs above, figure 14 and 15, analyze the pattern of traffic behavior in Vienna from Monday to Friday, therefore during the working week.

As shown in figure 16, cluster 3 includes the most counting points and is widely distributed across the entire city, including both central and suburban regions. This suggests that cluster 3 represents the standard or baseline traffic pattern for motor vehicles in Vienna. These locations likely experience moderate and consistent traffic volumes, typical for urban areas with balanced infrastructure and accessibility.

Cluster 2 appears more geographically concentrated, especially in central and northern parts of the city. The clustering of these points indicates areas with higher traffic volumes or specific urban characteristics, such as fewer alternative transport options for example in the 18th district or denser commercial zones. Particularly, many of these points are located outside the innermost districts but not in the far periphery, probably representing areas with limited subway access, where cars and trams serve as primary transportation modes.

Cluster 1, the smallest group, includes only a few stations that are mostly situated in the city center, for example "Karlsplatz" or counting points near the "Gürtel", with one permanent counting point known as "Wientalstraße" in the 14th district, which is in the outskirts of Vienna. These outlier locations might experience higher traffic volumes compared to the rest of the city, suggesting that they are areas near major roads or highways where traffic levels peak due to regional inflows and outflows or in the city center due to commuters.

In comparison to figure 15, the shortened time horizon does not affect the traffic behavior of Viennese people.

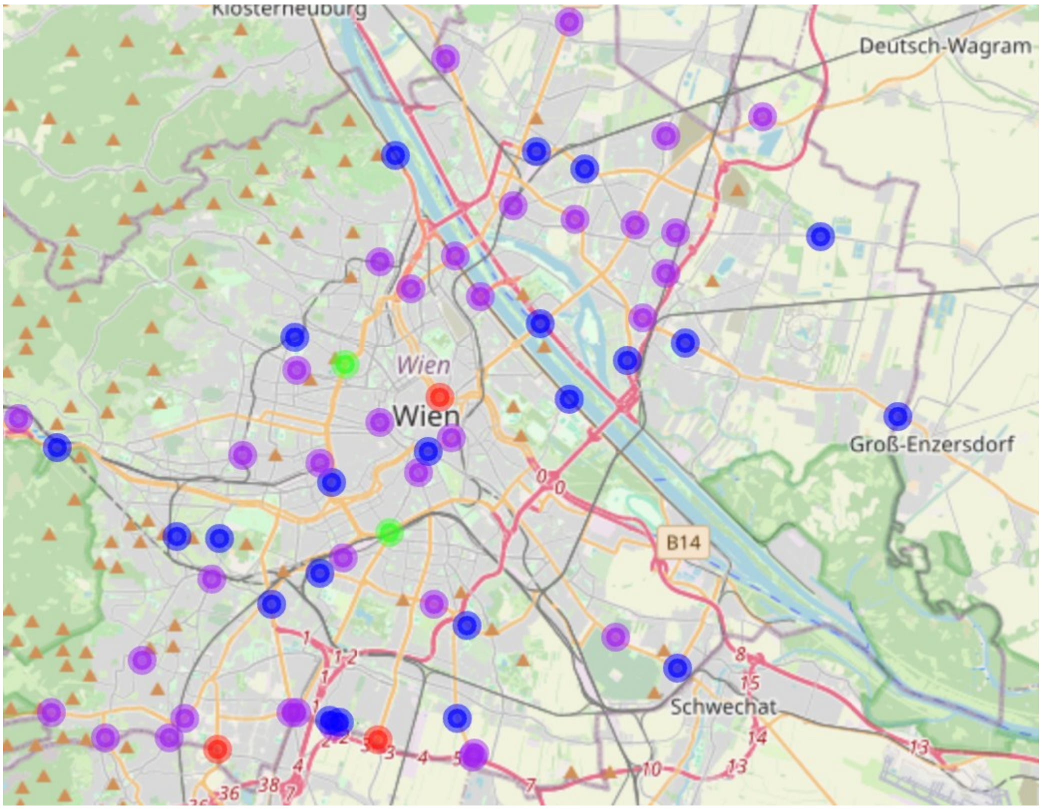

Figure 17 Cluster Analysis Lkw between Monday and Friday – Cluster 1: red, Cluster 2: green, Cluster 3: blue, Cluster 4: purple

The cluster analysis of heavy vehicle traffic (Lkw) at Vienna's permanent counting stations, as shown in figure 17, reveals four distinct traffic behavior patterns across the city.

Cluster 4 is the largest and most widely spread across Vienna. It contains both central and peripheral districts, suggesting that it showcases the standard or average traffic pattern for heavy vehicles. Cluster 3 also includes a significant number of stations but shows a slightly different spatial concentration. Many of its points are located in the eastern and southeastern areas of Vienna. This distribution could indicate regions with more active commercial or industrial activity, where truck volumes are somewhat elevated due to the presence of logistics centers or major distribution routes. In contrast, cluster 1 and cluster 2 are much smaller and more dispersed. These clusters represent outliers with unique traffic patterns. The red cluster may highlight locations with exceptionally high truck traffic, possibly situated near highway entry or exit points, freight terminals, or key transport corridors. The green cluster might include stations with low or irregular traffic volumes, or areas where truck access is limited or highly regulated.

Overall, the introduction of a fourth cluster allows for a more refined analysis of traffic dynamics throughout the city. It reveals not just general trends but also site-specific deviations that are critical for informed urban planning.

8.3. Clustering Analysis – Weekend and Holidays

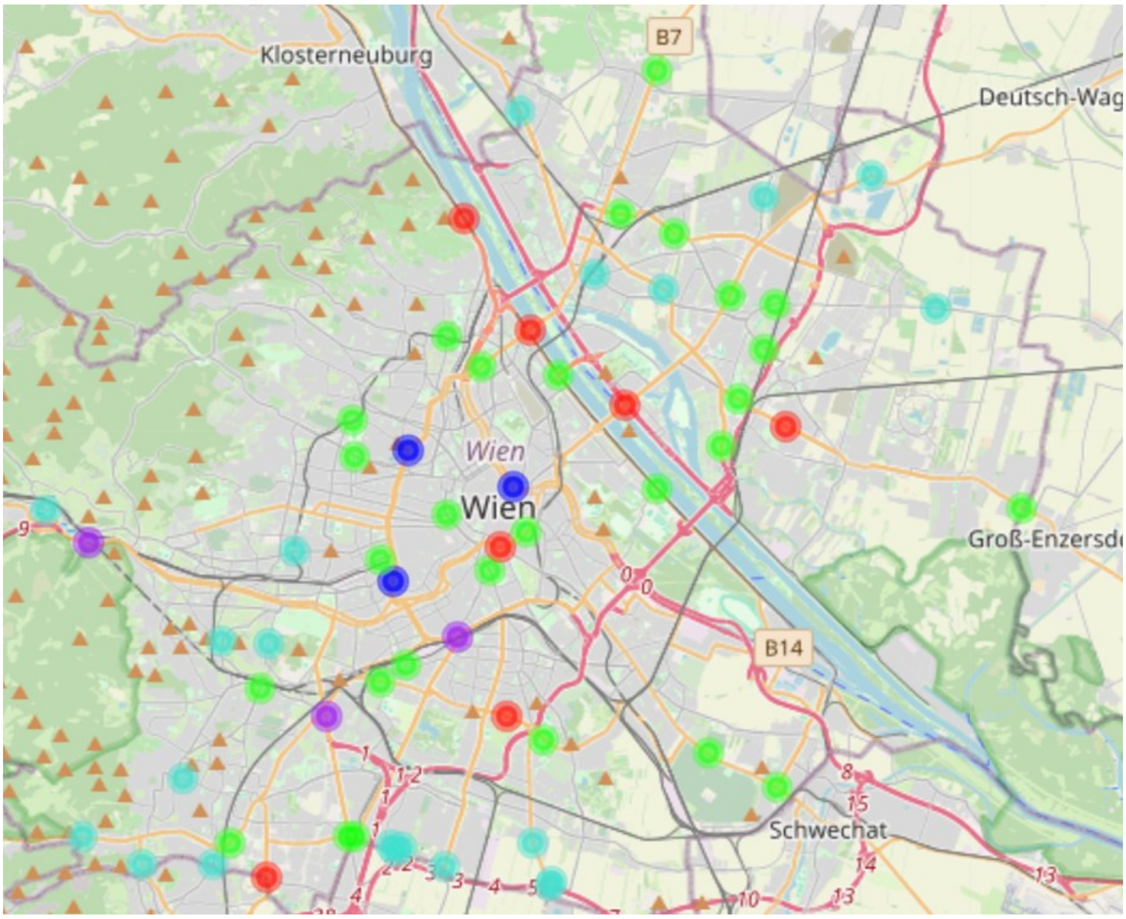

Figure 18 Cluster Analysis Kfz Weekends and Holidays – Cluster 1: red, Cluster 2: green, Cluster 3: turquoise, Cluster 4: blue, Cluster: purple

The following section analyzes the traffic behavior in Vienna during the weekends.

Figure 18 shows the result of a five-cluster analysis of motor vehicle traffic data from permanent counting stations across Vienna. By applying this more detailed clustering, the graph reveals more precise patterns and regional differences in traffic intensity and behavior throughout the city.

Cluster 2 is the most dominant and widespread group in the dataset. It spans across central and northern parts of Vienna, suggesting that it characterizes areas with typical or moderate traffic volumes. These are likely main urban arteries or mixed-use roads that experience consistent and balanced traffic flow without major anomalies. Cluster 3 is also well-represented, particularly in the southern and eastern parts of the city. This group could reflect roads with slightly higher traffic volumes, possibly connecting outer districts with commercial zones or acting as transit routes between districts without entering the city center directly.

Cluster 1 shows more selective distributions. Red points are distributed evenly across the city, indicating diverse but slightly deviating traffic patterns that don't fit neatly into the larger clusters. Cluster 5 is the smallest group and clearly the most distinct with only two locations. It is scattered throughout the southern and southwestern regions and likely identifies locations with highly specific traffic behaviors such as "Wientalstraße" and "Altmannsdorfer Straße". These areas are closely connected with highway exit points. An explanation for this high traffic volume can be found in the fact that a lot of Viennese people tend to leave the city on the weekends and are therefore measured by these counting points.

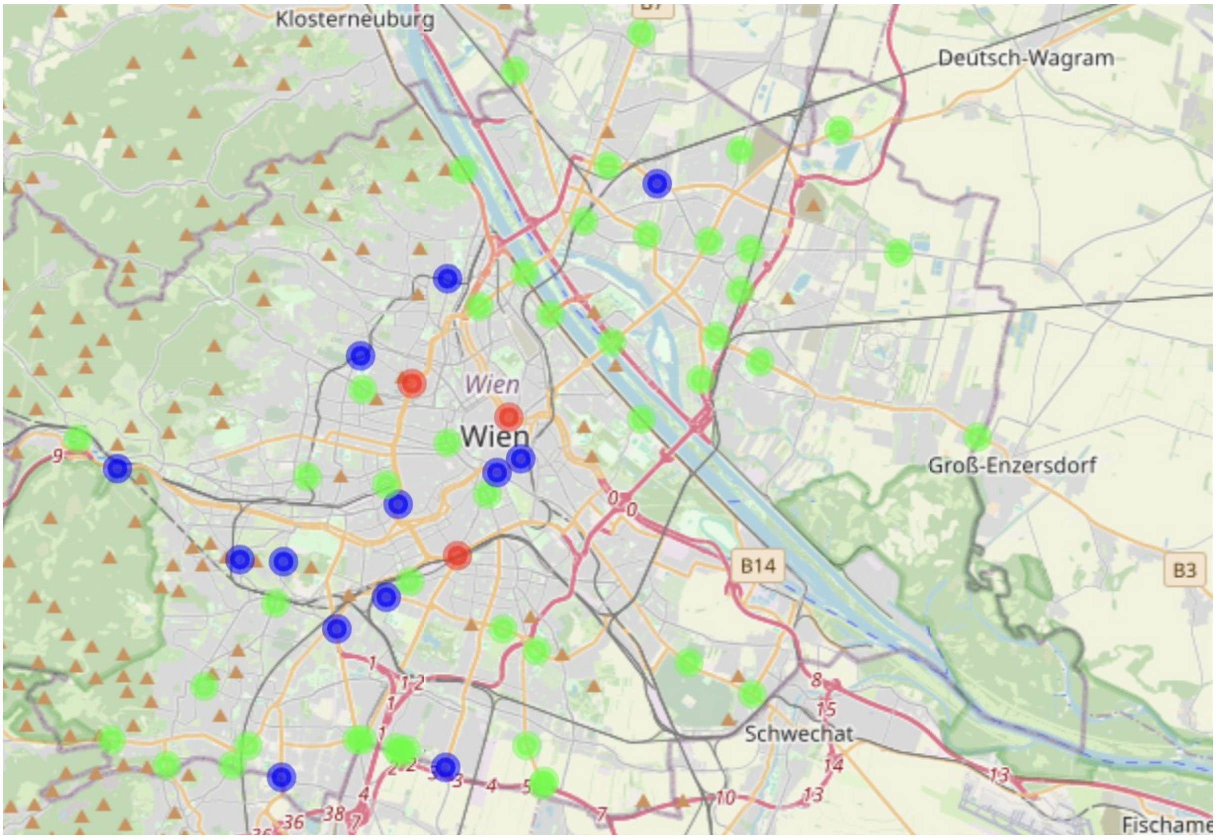

Figure 19 Cluster Analysis LKW Weekend and Holidays – Cluster 1: red, Cluster 2: green, Cluster 3: blue

We are focusing now on the graph titled figure 19. It illustrates a clustering of Vienna's permanent traffic counting stations based on heavy vehicle (LKW) data on the weekends.

Cluster 2 is the largest and most dominant group, covering the majority of counting stations throughout Vienna. Its widespread geographic presence, spanning both inner and outer districts, suggests that this cluster represents a standard or typical pattern of heavy vehicle traffic during this specific timeframe. These stations likely experience less freight activity, consistent with general urban logistics and transportation infrastructure.

Cluster 3 is much smaller and appears sporadically in the central and southern parts of the city. The limited number of stations in this group may reflect unique traffic conditions—perhaps slightly higher or irregular truck volumes due to local factors such as restricted access zones, public transport priority areas, or residential character.

Cluster 1 contains only a single station, indicating a strong outlier in the data. This lonely point may represent an exceptionally high or low volume of heavy vehicles, or a location with a highly atypical traffic pattern.

Overall, the clustering reveals that while much of Vienna shares a common traffic pattern, there are clear local variations. These patterns may reflect differences in urban density, public transport coverage, commercial activity, or road infrastructure. Understanding these clusters can help inform traffic planning and policy, particularly when considering infrastructure development, traffic regulation, or public transportation enhancements.

9. Regression Analysis Public Transportation

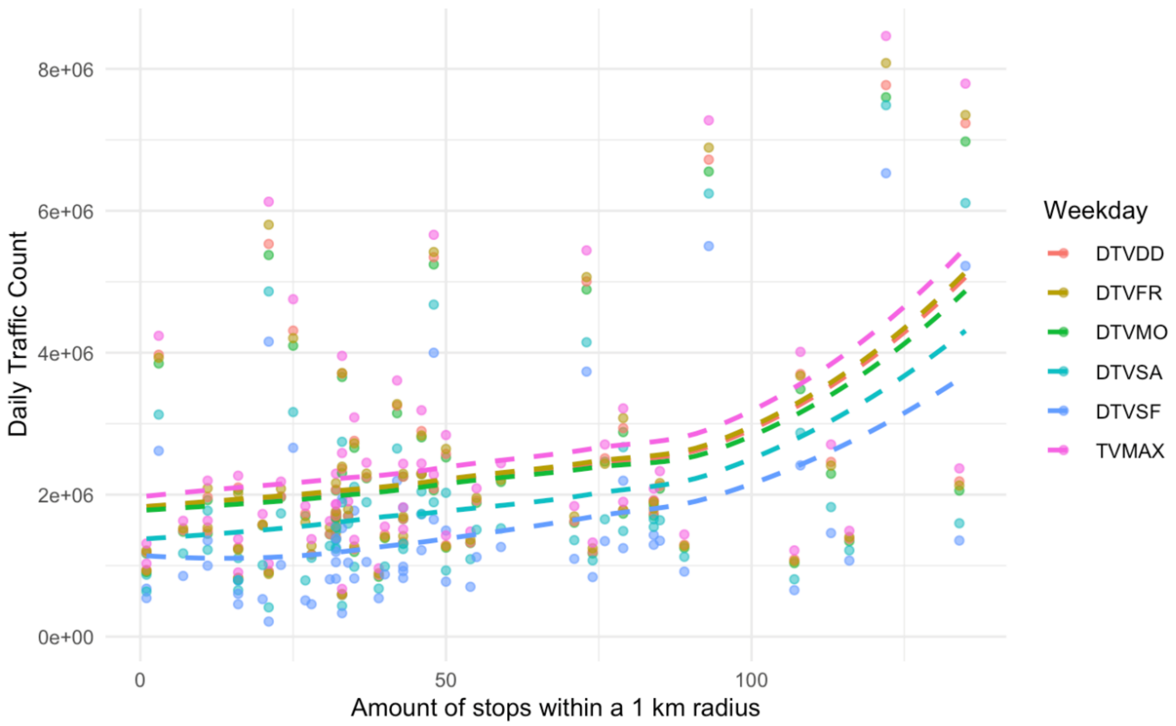

Figure 20: Correlation between Permanent Counting Stop Counts and Public Transportation – x-axis displays the amount of public transportation stops within a 1km radius and the y-axis displays the traffic volume

Figure 20 illustrates the relationship between the number of public transportation stops within a 1 km radius and the corresponding daily traffic volume. Each colored line in the graph represents data for specific days of the week, such as Monday (DTVMO), Tuesday to Thursday (DTVDD), Friday (DTVFR), Saturday (DTVSA), Sunday and holidays (DTVSF), and the maximum traffic value observed during the week (TVMAX).

On the x-axis, the number of public transport stops is plotted, while the y-axis shows the daily traffic volume. A general upward trend is visible across all categories, indicating that areas with a higher number of transits stops tend to experience greater daily traffic volumes. This positive correlation suggests that regions with better public transport infrastructure are typically more active, possibly due to higher urban density or due to the fact that people who live in the suburbs tend to drive to their nearest public transportation transit stop and then take public transportation for the following commute.

Interestingly, the trend lines show a nonlinear pattern. While traffic volume increases steadily with the number of stops, this increase becomes more pronounced once the number of stops exceeds approximately 80 to 100. This inflection point signals a threshold where public transport density begins to have a significantly stronger impact on traffic volumes.

The chart also highlights differences between weekdays and weekends. Traffic levels on weekdays, especially Monday to Friday, are generally higher, with closely aligned trend lines. In contrast, Saturday (DTVSA) and Sunday (DTVSF) generate lower daily traffic values, with Sunday showing the lowest levels overall. As expected, the line representing the weekly maximum (TVMAX) consistently lies above all others.

Despite the overall trend, some outliers are visible. A few locations with relatively few stops still report high traffic volumes. These may be special cases, such as areas dominated by car traffic or major roadways with limited public transport infrastructure.

Overall, the graph underscores the strong link between public transport density and urban traffic levels. For urban planners and policymakers, these insights are extremely valuable. They suggest that investing in dense and accessible public transport networks can influence traffic behavior, support sustainable mobility, and inform decisions about zoning, infrastructure upgrades, and traffic management strategies.

9.1. Regression Kfz

The following section of this report examines the relationship between the number of permanent counting points and the variation in traffic patterns across different weekdays and vehicle types. Using simple linear regressions this section will analyze whether a statistical significant relationship can be found.

The following three regressions focus solely on the vehicle type “Kfz”.

9.1.1. Regression Kfz Mo – Fr

Regression Formula: $\text{count}_i = \beta_0 + \beta_1 \cdot \text{DTVMO} + \beta_2 \cdot \text{DTVDD} + \beta_3 \cdot \text{DTVFR} + \varepsilon_i$

Coefficients:

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4422767	984719.582	4.491	3.28e-05 ***
count	50736	16330	3.107	0.00289 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure 21: Regression Kfz Monday - Friday

The regression examines the relationship between the number of public transportation and the daily traffic volume between Monday and Friday. The model includes a single predictor variable, count, which captures the sum of public transportation stops that are within a radius of 1km of a permanent counting point.

The result indicates that the regression intercept is approximately 4 422 767, with a standard error of 984 719.6. This suggests that when the number public transportation near counting points is zero, the predicted traffic volume is about 4.42 million. The intercept is highly statistically significant, with a p-value of 3.28e-05, which is well below the common significance threshold of 1%.

The coefficient for the variable count is estimated at 50 736 with a standard error of 16 330. This means that for each additional public transport stop near permanent counting point, the model predicts an increase of approximately 50 736 in daily traffic volume. The t-value for this coefficient is 3.107, and the corresponding p-value is 0.00289. This result is statistically significant at the 1% level, indicating strong evidence of a positive relationship between the number of counting points and traffic volume.

Overall, the model suggests a statistically and practically significant connection: areas with more public transport infrastructure tend to experience higher traffic volumes. This finding supports the idea that the density of traffic measurement points is positively associated with actual traffic activity, possibly reflecting more urbanized or heavily used transport networks.

9.1.2. **Regression Kfz Weekend and Holidays**

The next regression focusses on the relationship between the number of public transportation stops and the traffic $\text{count}_i = \beta_0 + \beta_1 \cdot \text{DTVFR} + \beta_2 \cdot \text{DTVSF} + \varepsilon_i$ volume during weekends and on holidays.

Regression Formula:

Coefficients:

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1792480	559849	3.202	0.00219 **
count	32056	9284	3.453	0.00102 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure 22: Regression Kfz Weekends and Holidays

This regression also supports the idea that the traffic volume increases as the number of public transportation increases. This positive correlation predicts that the traffic volume increases by 32 056 if the one public transportation stop is added within a 1km radius of a permanent traffic count. With a p-value of 0.00102 this relationship is highly significant.

In comparison to the traffic behavior during the week the pattern does not change. Although the traffic volume decreases drastically for weekends, the positive relationship still remains.

9.1.3. **Regression Kfz MAX**

Regression Formula: $\text{count}_i = \beta_0 + \beta_1 \cdot \text{DTVMAX} + \varepsilon_i$

Coefficients:

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1625808	354780	4.583	2.38e-05 ***
count	18377	5883	3.124	0.00275 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 23: Regression Kfz maximum value

Following the regressions above, this regression shows that the maximum value per day is significantly dependent on the amount of stops within a 1km radius. The table shows that the traffic volume increases by 18 377, if one additional public transportation stop is added.

9.2. Regression LKW

The following three regressions focus solely on the vehicle type "LKW".

9.2.1. **Regression LKW Mo – Fr**

Regression Formula: $\text{count}_i = \beta_0 + \beta_1 \cdot \text{DTVMO} + \beta_2 \cdot \text{DTVDD} + \beta_3 \cdot \text{DTVFR} + \varepsilon_i$

Coefficients:

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	355355.8	73314.4	4.847	9.23e-06 ***
count	920.9	1215.8	0.757	0.452

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 24: Regression Lkw Monday - Friday

This regression analyzes the relationship between public transportation stops and the traffic volume generated by heavy vehicles. The result indicates that no significant relationship between public transportation exists due to a p-value of 0.452. This result is intuitive since truck drivers do not depend on public transportation infrastructure to go to work for example. Therefore, heavy vehicle traffic does not be considered when planning future public transportation infrastructure.

9.2.2. **Regression LKW MAX**

Regression Formula: $\text{count}_i = \beta_0 + \beta_1 \cdot \text{DTVMAX} + \varepsilon_i$

Coefficients:

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	138291.2	27669.8	4.998	5.33e-06 ***
count	356.5	458.9	0.777	0.44

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 25: Regression Lkw maximum value

This regression follows the same intuitive explaination of the result as above, since the p-value is not significant when analyzing the maximum traffic value.

10. Conclusion and Results

This project successfully conducted a multi-faceted analysis of Vienna's traffic patterns, leveraging historical data from 2016 to develop a deep understanding of urban mobility and create robust forecasts extending to 2030. By integrating descriptive, spatial and predictive analytics, the project has yielded significant insights and produced a powerful interactive tool for strategic decision-making.

10.1. Key Findings and Results

Throughout the various stages of analysis, several key findings emerged that characterize the complex dynamics of traffic in Vienna:

1. Distinct Urban and Suburban Traffic Profiles: A clear divergence in traffic trends between Vienna's city center, comprising districts 1 through 9, and its suburbs was identified. The city center exhibits higher traffic volumes with more pronounced seasonal fluctuations and a significant, though recovering, impact from the COVID-19 pandemic. Suburban traffic, while lower in volume, remained more stable and less affected by the pandemic's disruptions, indicating different underlying mobility drivers.

2. Resilience of Commercial vs. Passenger Traffic: The COVID-19 analysis revealed the starkly different impacts of the pandemic on passenger cars, or Kfz, versus trucks, known as Lkw. Passenger car volume saw a dramatic and immediate decline during lockdown periods, reflecting shifts in commuting and personal travel. In contrast, truck traffic demonstrated remarkable resilience, experiencing only minor dips and highlighting its essential role in maintaining supply chains, even during a public health crisis.

3. Spatial Distribution and Hotspots: The heatmap analysis provided a clear spatial visualization of traffic density. Passenger car traffic is heavily concentrated in the city center and along major commuter arteries. Truck traffic, however, follows logistical and industrial corridors, with higher densities observed in peripheral areas and near major highways. The combined view reveals critical congestion zones where both traffic types overlap, particularly on ring roads and key transport nodes.

4. Behavioral Clustering of Counting Stations: The k-means clustering analysis successfully grouped counting stations with similar traffic behaviors, uncovering patterns that were not immediately obvious. It differentiated between weekday commuter traffic and weekend leisure traffic, and effectively isolated outlier stations, such as those near major highway access points like Wientalstraße, which exhibit inner-city traffic volumes despite their suburban location. This demonstrates that geographic location alone is not the sole determinant of traffic behavior.

5. Impact of Public Transportation Infrastructure: The regression analysis established a statistically significant positive correlation between the density of public transport stops and the volume of passenger car traffic. This suggests that areas well-served by public transport are also hubs of high activity and mobility. Conversely, no significant relationship was found for truck traffic, confirming that freight logistics operate independently of the public transit network.

10.2. Conclusion and Outlook

This project has successfully demonstrated the value of a holistic data science approach to understanding urban mobility. By moving beyond simple traffic counts to incorporate spatial analysis, predictive modeling, and the influence of external factors, we have created a nuanced and actionable picture of Vienna's traffic network.

The development of a weighted ensemble forecast, which leverages the strengths of SARIMA, Prophet, and SARIMAX models, has resulted in a robust and reliable predictive tool. The superior performance of the SARIMA model in our backtesting highlights the strong influence of historical patterns in predicting future traffic, while the inclusion of Prophet and SARIMAX provides a safeguard against model-specific weaknesses.

The culmination of this work is the Vienna Car Traffic Forecast Dashboard. This interactive tool translates complex datasets and model outputs into an intuitive and accessible format. It empowers stakeholders at KPMG and in urban planning to not only view historical trends but also to explore future scenarios, compare the projections of different models, and understand the factors driving traffic changes at a granular, station-by-station level.

Looking forward, this project lays a strong foundation for further research. Future work could involve incorporating more granular data, such as real-time weather conditions, specific public events, or road construction schedules, to further refine the SARIMAX model's predictive accuracy. Additionally, the clustering analysis could be extended to identify optimal locations for new infrastructure or to model the impact of new mobility solutions.

In conclusion, this project provides not just a forecast, but a comprehensive analytical framework and a powerful visualization tool that can support strategic planning, optimize traffic management, and contribute to the development of a more efficient and resilient transportation system for the city of Vienna.

11. Literature list

Dieckmann, J. (2024, May 28). Getting Started Predicting Time Series Data with Facebook Prophet. Medium. https://medium.com/data-science/getting-started-predicting-time-series-data-with-facebook-prophet-c74ad3040525

12.2 Prophet model | Forecasting: Principles and Practice (3rd ed). (n.d.). https://otexts.com/fpp3/prophet.html

Rasulmukhamedov, M., Tashmetov, T., & Tashmetov, K. (2024). Forecasting Traffic Flow Using Machine Learning Algorithms. *EEPES 2024*, 14. https://doi.org/10.3390/engproc2024070014

Shumway, R.H., Stoffer, D.S. (2017). ARIMA Models. In: Time Series Analysis and Its Applications. Springer Texts in Statistics. Springer, Cham. https://doi.org/10.1007/978-3-319-52452-8_3

Zhang, Y. N., & Haghani, A. H. (2015). A gradient boosting method to improve travel time prediction. In *Department of Civil & Environmental Engineering, University of Maryland, College Park, MD 20740, United States* (MD 20740). Elsevier. https://www.sciencedirect.com/science/article/abs/pii/S0968090X15000741

Xia, Y. X., & Chen, J. (2017). *Traffic flow Forecasting method based on gradient boosting Decision tree* (CC BY-NC 4.0). https://doi.org/10.2991/fmsmt-17.2017.87

A review of OpenStreetMap data; P Mooney, M Minghini - Mapping and the citizen sensor, 2017 - library.oapen.org

EPSG Registry, https://epsg.io/31256

12. List of figures