Lucerne University of Applied Sciences and Arts (HSLU)

Master of Science in Applied Information and Data Science
Data Warehouse and Data Lake

Project Proposal Data Warehouse and Data Lake

The Influence of Weather and Demographic Parameters on the Shared Mobility Usage in Switzerland

Authors: Eduardo Ramiro, Pacheco Ardon, 21-864-806, eduardo.pachecoardon@stud.hslu.ch

Raphael Suter, 17-475-724, raphael.suter.01@stud.hslu.ch

Nithusya Mohanadasan, 14-702-864, nithusya.mohanadasan@stud.hslu.ch

Examiners: PD Dr. Luis Terán, José Mancera

Place and date: Lucerne, 12.03.2023

Table of Contents

1 Background and Motivation

In the recent years, the concept of a “sharing economy” has risen in fame and led to an immense success of digital platforms such as Airbnb or eBay. The idea is based on the paradigm that for example products, services and personal skills are used across ownerships and hence lead to a better distribution of underutilized assets and sustainable consumption in the society [Machado et al. 2018]. One such implementation is called “Shared Mobility” or “Mobility as a Service” (MaaS) [Bundesamt für Energie BFE 2023]. A lack of urban planning and the increasing number of vehicles in the cities not only cause traffic jam problems but also air pollution and therefore lead to a worsening of the life quality of citizens [Machado et al. 2018]. Shared mobility can be described as a “short-term access to shared vehicles according to the user’s needs and convenience” [Machado et al. 2018, p. 1]. By enforcing sustainable travel behavior using existing urban infrastructure and meeting social needs of equity, the quality of life is improved again [Shokouhyar et al. 2021].

Also, in Switzerland this strategy is well-known and used. According to EnergieSchweiz (2023), the government’s program for energy efficiency and renewable energy, all categories in shared mobility have increased strongly in the past years. The main categories include car-sharing, bike-sharing, and E-Scooter-sharing. All models follow the same concept, which is using a digital platform the availability of a vehicle is verified and then booked for a ride. After the use, the vehicle is brought back to the original place or to a dedicated parking space. In Switzerland around 6’000 cars, 11’000 bikes and 6’000 E-Scooters are shared for this purpose [EnergieSchweiz 2023].

The so-called “Shared Mobility Agenda 2030” (SMA 2030), presented by EnergieSchweiz, depicts the current shared-mobility market, and defines the vision for Switzerland. It says that the goal is that shared mobility should become an “integrated part” in the Swiss transport system “protecting the natural environment, ensuring economic efficiency and ensuring social solidarity” [Schmid et al., p.2, own translation]. The goal of this project is to examine the shared mobility data by building a data lake and data warehouse using AWS technologies, to provide custom analysis for relevant decision-makers.

2 Problem Statement and Research Questions

As discussed above, many benefits exist for the public and government by using/offering shared mobility services. However, many shared mobility systems providers face problems regarding optimal resource allocation as Laporte et al. (2015) point out. How many vehicles should be maintained at a station, such that the demand at this place is met? When and how should vehicles be moved to low-inventory stations? Spaces located close to the airport or railway stations are very efficient. Nevertheless, also less frequented locations should be encouraged to use these systems. Therefore, important aspects and decisions must be considered by providers [Laporte et al. 2015].

One such aspect or influencing factors are weather conditions. It is known that weather conditions and daily human mobility patterns are linked [Kimpton et al. 2022]. One expected effect of climate change will be an increase in intra- and inter-seasonal variations in various parts of Europe. These extremes will disproportionately affect cities since these host most population and economic activity and are often located in climate sensitive areas such as plains and coastal zones [Heyndrickx et al. 2014].

The literature suggests that especially bike-sharing and E-Scooter sharing are affected by the weather. Unfavorable conditions, such as rain or snow, usually discourage cycling and decrease the number E-Scooter trips and its distance [Machado et al. 2018 & Kimpton et al. 2022]. In this sense, providers need to adjust their offers across seasons. We would like to further explore this relationship between weather and the usage of shared mobility services in Switzerland across the distinct categories. Moreover, demographics data from the cities/cantons will be added, to capture any demographic influence in the usage of shared mobility. Thus, our research questions are:

1. Which parameters (related to demographics and weather) in the selected time period, show a significant correlation with the usage of shared mobility services, at 95% confidence level?

2. Which parameters (related to demographics and weather) in the selected time period, show a significant correlation with the usage of shared mobility services across categories (car, bike, E-Scooter), at 95% confidence level?

With question one and two, we pretend to identify the most relevant variables that influence the shared mobility demand and want to provide stakeholders with important insights. These parameters should be focused on in a further analysis by stakeholders. Here we can also identify which data source has a greater influence in predicting the share mobility demand.

3. How accurate would be a demand forecast for the shared mobility services based on the fetched data, by city/canton for one week ahead, compared to the real data?

With the last research question, we pretend to provide the stakeholders with a visualization forecast in Tableau, to quantify in advance an approximated demand of vehicles in the selected location, so that vehicles and resources can be efficiently allocated.

To answer these questions, shared mobility data is extracted using an API and loaded into the data lake multiple times a day. Then climate data for these times would be extracted from a weather API, such that changing weather conditions can be captured. The usage of offers is quantified by the difference in numbers of available vehicles at a station across the extraction times e.g., if a station has 2 cars available at 8 am and then only one car available at 10 am, this means 1 car has been used (usage = 1). This variable will then be tracked during the day and across weather conditions.

3 Data Sources

3.1 Shared Mobility API (Dynamic):

This data source, provided by opendata.swiss, shows the locations of stations and the availability of shared mobility vehicles in real-time [opendata.swiss 2023].

Website: https://opendata.swiss/de/dataset/standorte-und-verfugbarkeit-von-shared-mobility-angeboten
Base URL: api.sharedmobility.ch/v1/sharedmobility

In the response a JSON file can be found with geolocation data:

x	0
y	0
spatialReference	{'wkid': 4326}

And attributes data:

provider.id	string
provider.name	string
provider.timezone	string
provider.apps.ios.store_uri	string
provider.apps.android.store_uri	string
id	string
pickup_type	free_floating
station.name	string
station.address	string
station.postcode	string
station.status.installed	true
station.status.renting	true
station.status.returning	true
station.status.num_vehicle_available	0
station.status.num_vehicle_disabled	0
vehicle.status.disabled	true
vehicle.status.reserved	true
vehicle_type	string

For this project, the attributes “x” and “y” (geometry coordinates of the location), “station.name” (name of the station), “station.status.num_vehicle_available” (number of available vehicles) and “vehicle_type” are the most relevant variables.

3.2 Weather API (Dynamic):

SRF Meteo serves weather forecast for over 100'000 locations within Switzerland. After providing either of name, zip code or decimal latitude and longitude parameters of the chosen location, the API sends a response that contains a JSON object based on the request with the weather forecast of the location for the chosen time period (hourly, daily) [SRG SSR 2023].

Website: https://developer.srgssr.ch/api-catalog/srf-weather
Base URL: api.srgssr.ch/srf-meteo

In the response a JSON file can be found with geolocation data:

id	0
lat	0
lon	0
station_id	string
timezone	string
default_name	string
alarm_region_id	string
alarm_region_name	string
district	string
geolocation_names	[{'district': 'string', 'id': 'string', 'locat...

And forecast data:

local_date_time	string
TTT_C	string
TTL_C	0
TTH_C	0
PROBPCP_PERCENT	string
RRR_MM	string
FF_KMH	string
FX_KMH	string
DD_DEG	string
SYMBOL_CODE	0
type	0
cur_color.temperature	0
cur_color.background_color	string
cur_color.text_color	string

3.3 Demographics API (Static):

This dataset provides information of the permanent and non-permanent resident population regarding institutional units, citizenship (category), sex and age [Federal Statistical Office 2022].

The data will be extracted from an interactive table in csv format. Unfortunately, there are no data available yet for the year 2022. The data available is from year 2010 up to 2021, therefore we will be using the demographics from year 2021.

Switzerland's permanent resident population was 8 738 800 at the end of 2021, i.e., 0.8% more than in 2020. In this population, one woman in five and one man in six were aged over 64. The country had 162 more centenarians than in the previous year. These are the final results of the population and households' statistics from the Federal Statistical Office (FSO) [Federal Statistical Office Reports 2022].

Variable	**Description**
Year	Year of the data from 2010 up to 2021
Canton (-) / District (>>) / Commune (.....)	It is possible to get the data on canton, district or commune level.
Population type	It is possible to choose between "permanent resident population" and "non permanent resident population"
Citizenship (category)	There are two different categories: Switzerland and foreign country
Sex	There are two different categories: Male and Female
Age	#citizens of the age from 0 years to 100 years or older

Website:

https://www.bfs.admin.ch/bfs/en/home/statistics/population/effectif-change.assetdetail.23044769.html

4 Objectives and Output

In general, the goal of our project is first to get familiar with the AWS technologies and learn how to work with them and real-time data. However, we can identify other technologies and technical tasks that we will aim to learn and practice as well:

- To design and implement data pipelines, which will fetch the required data in real time.
- To implement and get familiar with Apache AirFlow, and to contrast the differences with the AWS Lambda function.
- To understand the concepts and main differences between a Data Lake and a Data Warehouse, while implementing such concepts on a real project.

- To get familiarized and use PostgreSQL to query some data subsets, which should be used to produce the required visualizations.
- To create a Github repository and push changes on the files through Git in a proper manner.

Next to that we also want to make an analysis of the current situation of mobility rentals in Switzerland. In this regard we have defined the following goals:

- To identify which data source (demographics or weather data) and which variables, have the higher influence on the mobility services demand.
- To find out the regions of Switzerland and which type of rentals (E-bike, E-Scooter, cars) have a higher demand.
- To produce an accurate forecast system for the interested stakeholders and companies, as a tool which would provide the guidelines to efficiently allocate resources.

5 Addressee

Our findings are beneficial for any company, which provides shared mobility services in Switzerland. For this reason, the result of our work could be provided to those companies to expand their business or just to improve the current business model. With that, the revenue can be increased as our analysis helps them to better understand where and at which weather conditions, the vehicles will be used most/least. This can lead to an increase in their turnover. Furthermore, with the right scheduling of the collection of the E-scooters for charging, the company can make sure that all their E-scooters can be used when they are needed. If the demand for E-scooters is high but they cannot be rented out, as they ran out of battery, the company loses money, and the customer will be unsatisfied.

6 Proceeding and Method

For our project we follow the phases of the CRISP-DM (Cross-Industry Standard Process for Data Mining) model. It is one of the most popular process models for data mining and analytics projects [Data Science Process Alliance 2022]. The CRISP-DM model provides a structured approach to the entire data mining process, from understanding the business problem to deploying the model. The CRISP-DM model consists of the following six phases [Chapman et al. 2000]:

1. Business Understanding: In this phase, we define the objectives of the project and try to understand the business problem. This phase we covered already by writing this project proposal.
2. Data Understanding: This phase involves gathering and understanding the data that will be used for the analysis. As explained in chapter 2 we will gather the data via API using the lambda function from AWS. Furthermore, in this phase the test account will be used to determine the number of extractions per day needed to provide a meaningful analysis and how long the selected period should be. As of now, a period of 2-3 weeks is planned.
3. Data Preparation: In this phase, the data is cleaned, transformed, and formatted in preparation for the analysis.
4. Modeling: This phase involves the development of predictive models using the prepared data.
5. Evaluation: In this phase we analyse our data and answer our research questions.
6. Deployment: In the final phase, the developed models are deployed into production to solve the business problem

The CRISP-DM process is an iterative one, meaning that it is designed to be flexible and allow for adjustments to be made throughout the project based on feedback and insights gained from each phase of the process. Unlike the traditional waterfall model, the order of the phases in CRISP-DM is not fixed and moving back and forth between phases is often necessary. As a result, the model indirectly promotes the principles and practices of agility.

7 Time Schedule

The time schedule we plan to follow is presented below:

When	What	Explanation
13.03.23 - 26.03.23	Data Understanding	- Understand the data - Retrieve data from APIs - Test the framework in AWS
27.03.23 - 02.04.23	Set up Data Ingestion and Data Storage in AWS	- Set up Data Pipelines - Set up ETL Pipelines - Amazon Lambda Functions - Set up Storage in S3
03.04.23 - 23.04.23	Collect Data in Data Lake	- Design and implementation
06.04.23	Mid Term Presentation	- Showcase advances and standing point
10.04.23 - 23.04.23	Data Cleaning and Data Transformation in AWS	- Before setting up the Data Warehouse the data needs to be prepared
24.04.23 - 30.04.23	Set up Data Warehouse	- Schema integration - Design data warehouse structure
01.05.23 - 14.05.23	Data Analysis	- Answer the business questions - Use predictive models for forecasting
15.05.23 - 21.05.23	Data Visualization in Tableau	- Produce insightful visualizations - Use graphical forecast
25.05.23	Final Presentation	- Present findings and results

8 Literature

[Bundesamt für Energie BFE 2023] Bundesamt für Energie BFE: *Standorte und Verfügbarkeit von Shared Mobility Angeboten*. available: https://opendata.swiss/de/dataset/standorte-und-verfugbarkeit-von-shared-mobility-angeboten, accessed 05th March 2023.

[Chapman et al. 2000] Chapman, Pete; Clinton, Julian; Kerber, Randy; Khabaza, Thomas; Reinartz, Thomas; Shearer, Colin; and Wirth, Rüdiger. *CRISP-DM 1.0: Step-by-step data mining guide*. available: https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf, accessed 5th March 2023

[Data Science Process Alliance 2022] Data Science Process Alliance: *CRISP-DM is Still the Most Popular Framework for Executing Data Science Projects.* available: CRISP-DM is Still the Most Popular Framework for Executing Data Science Projects - Data Science Process Alliance (datascience-pm.com), accessed 11th March 2023.

[EnergieSchweiz 2023] EnergieSchweiz: *Shared Mobility: Fahrzeuge teilen statt besitzen*. available: https://www.energieschweiz.ch/mobilitaet/shared-mobility/, accessed 05th March 2023.

[Federal Statistical Office 2022] Federal Statistical Office: *Permanent and non-permanent resident population by institutional units, citizenship (category), sex and age.* available: https://www.bfs.admin.ch/bfs/en/home/statistics/population/effectif-change.assetdetail.23044769.html, accessed 11th March 2023.

[Federal Statistical Office Reports 2022] Federal Statistical Office Reports: *Population size and change in Switzerland in 2021: final results 2022.* Available in DE, FR, IT: https://www.bfs.admin.ch/news/de/2022-0467, accessed 11th. March 2023.

[Kimpton et al. 2022] Kimpton, Anthony; Loginova, Julia; Pojani, Dorina; Bean, Richard; Sigler, Thomas; Corcoran, Jonathan: Weather to scoot? How weather shapes shared e-scooter ridership patterns. *Journal of Transport Geography,* Vol. 104 (2022), 103439.

[Heyndrickx et al. 2014] Heyndrickx, Christophe; Purwanto, Joko; Ciari, Francesco; Bösch, Patrick M.; Perrels, Adriaan: The impact of extreme weather events on urban mobility in Switzerland. *ETH Zurich Research Collection*: https://www.research-collection.ethz.ch/

[Laporte et al. 2015] Laporte, Gilbert; Meunier, Frédéric; Calvo, Roberto Wolfler: Shared mobility systems. *4or*, Vol. 13 (2015), p. 341-360.

[Machado et al. 2018] Machado, Cláudia A. Soares; de Salles Hue, Nicolas P. Marie; Berssaneti, Fernando Tobal; & Quintanilha, José Alberto: An overview of shared mobility. *Sustainability*, Vol. 10, Nr. 12 (2018), 4342.

[opendata.swiss 2023] opendata.swiss: *Standorte und Verfügbarkeit von Shared Mobility Angeboten*. available: https://opendata.swiss/de/dataset/standorte-und-verfugbarkeit-von-shared-mobility-angeboten, accessed 11th March 2023.

[Schmid et al. 2022] Schmid, Jonas; Halef, Mathias; Dr. Beckmann, Jörg: Shared Mobility Agenda 2030 Synthesebericht. EnergieSchweiz (2022).

[Shokouhyar et al. 2021] Shokouhyar, Sajjad; Shokoohyar, Sina; Sobhani, Anae; Gorizi, Amirsalar Jafari: Shared mobility in post-COVID era: New challenges and opportunities. *Sustainable Cities and Society*, Vol. 67 (2021), 102714.

[SRG SSR 2023] SRG SSR: *SRF Weather.* available: https://developer.srgssr.ch/api-catalog/srf-weather, accessed 11th March 2023.