

# Machine Learning and Sustainable Mobility: The Case of the University of Foggia (Italy)

**Giulio Mario Cappelletti**

University of Foggia

**Luca Grilli**

University of Foggia

**Carlo Russo** (✉ [carlo.russo@unifg.it](mailto:carlo.russo@unifg.it))

University of Foggia

**Domenico Santoro**

University of Bari Aldo Moro

---

## Research Article

**Keywords:** University, Sustainability, Transport Policy, Mobility Choices, Machine Learning, Emissions

**Posted Date:** October 27th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-963685/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Thanks to the development of increasingly sophisticated machine-learning techniques, it is possible to improve predictions of a certain phenomenon. In this paper, after having analyzed data relating to the mobility habits of University of Foggia (UniFG) community members and determined their emissions of pollutants, we applied machine-learning techniques to these data to estimate the quantities of pollutants (in a certain time period) produced by new subjects not present in the data sets, using very little information. In this way, we developed a method that the university could apply to inform new students about what their emissions of pollutants could be in the near future, through several easily obtainable features. This method could allow the UniFG Rectorate to improve its sustainable mobility policies by encouraging the use of methods that are as appropriate as possible to the users' needs. In addition, any public/private organization outside the academic environment can use the method, due to the need for little information.

## 1. Introduction

The problem of sustainable mobility choices is involving more and more public and private organizations. As defined by Banister (2008), implementing the sustainable-mobility paradigm requires intervention by a series of people who can develop policies capable of reducing car dependence. Sustainable mobility plays a fundamental role (Jeon and Amekudzi 2005) in community development, from economic, environmental, and social points of view. In particular, a sustainable transport system should guarantee efficient travel to all users and an ecological means of transport while maintaining certain levels of economy and public health (Weichenthal et al. 2014, Lopez-Arboleda, Sarmiento, and Cardenas 2020, de la Torre et al. 2021).

Decision-makers' choices recently have been oriented toward reducing pollutant emissions in the most congested urban areas by introducing green vehicles such as bicycles or scooters; however, these means of transport represent only a part of cities' environmental development (Simons et al. 2014). The decisions taken include the introduction of car-sharing systems, integrated transport systems, bike-sharing systems, bus prioritization, and free public transport (Becker, Ciari, and Axhausen 2017, Chakhtoura and Pojani 2016, Tafidis, Sdoukopoulos, and Pitsiava-Latinopoulou 2017, Suchanek and Szmelter-Jarosz 2019).

The objective of this paper is to develop a system for evaluating the mobility choices of community members (in this specific case, the members of the University of Foggia) based on the distance covered and the types of vehicles used. Such a system would allow each member to know whether or not he/she is respecting the objectives set by the community policy-maker and consequently to try to improve himself/herself to meet the established parameters. In the particular case in question, we develop a system based on machine-learning (ML) algorithms with which to evaluate how sustainable the means of transport are of UniFG members and to receive the evaluation directly through a green/red sticker on the management software used by the university.

The present paper is structured as follows. In the following section, the variables present in the data set collected from the academic community are described, and the most important features are analyzed through factor analysis and machine learning. Section 3 presents a possible application of these features to determine the emission levels of each UniFG member. In section 4, some conclusions are drawn.

## 1.1. Application of Machine Learning

Researchers recently have used ML algorithms to solve many problems, ranging from the industrial to the economic spheres. Within the world of sustainable mobility, several authors have proposed models whose objectives range from determining the shortest route with a means of transport to determining which means to choose based on certain variables (called features). For example, using a Random Forest, Zhou, Wang, and Li (2019) determined the seasonality in the choice of bike sharing, as compared to taxis, in Chicago. Basu and Ferreira (2020) used neural networks to identify the factors influencing the choice of a type of means of transport. Yang et al. (2020) proposed a deep learning system to improve the forecast of the demand for bikes in bike sharing systems. Using machine-learning algorithms, Tang, Liu, and Choudhury (2020) were able to predict passengers' bus stops. Migliore, Burgio, and di Giovanna (2014) proposed a system based on ML to optimize parking prices in Palermo. In a Support Vector Machine (SVM) study, Liang et al. (2019) predicted household travel modes based on various factors. Asensio et al. (2020) solved several issues using text-mining algorithms on reviews for electric car charging stations. Nandal, Mor, and Sood (2020) highlighted the role of neural networks in improving infrastructure deterioration. Finally, Hasan, Whyte, and Jassmi (2020) developed a system capable of reducing pollutant emissions from self-driving vehicles by 80%.

In this paper, we will use of ML algorithms to predict some of the missing data necessary to estimate the emissions of UniFG members. In particular, this method will allow us to exploit data that are easily obtainable through a very short questionnaire in order to estimate other missing data, which would be more difficult to obtain for each community member in practice.

## 2. Materials And Methods

### 2.1. Emissions description

The first step in understanding the levels of total annual atmospheric emissions among UniFG members is to consider the largest possible number of variables detected during the survey (Cappelletti et al. 2021). In particular, we will consider the following parameters:

- The distance in kilometers that the respondents claim to travel;
- The frequency at which the respondents go to their reference facility per weekly;
- Whether it is the hot or cold season, which affects both the number of weeks of activity and different modes of travel; and
- The mode of transport used.

Notably, of these four parameters considered, the season is of particular importance precisely when considering its implications on the other parameters. For this reason, we considered two partial data sets with only the data relating to each chosen season for calculating the seasonal kilometers. Furthermore, because these partial data sets were aimed at calculating emissions, the records of all the respondents who indicated that they used travel methods without significant emissions (walking, cycling, and riding an electric scooter) for each period considered were eliminated. The result of this operation was two data sets, “hot season km” and “cold season km”, which contain 2385 and 2517 records, respectively. Finally, for a calculation that considers the peculiarities of the different travel modes, it is necessary to obtain a single distinct kilometer value for each travel mode. The formula used to calculate the single kilometer values was as follows:

$$KM_{sm} = \sum_{i=1}^{n_{sm}} Da/r_i \cdot FR_i \cdot SET_i$$

where  $KM_{sm}$  is the total kilometers travelled in season  $s$  by those who adopted travel method  $m$ ,  $n_{sm}$  is the total number of respondents who adopted travel method  $m$  in season  $s$ ;  $Da/r$  is the round-trip distance travelled in kilometers to reach the university structure;  $FR$  is the weekly travel frequency in season  $s$ , and  $SET$  is the average number of weeks of lessons in season  $s$ . This procedure was carried out for 63 modes of travel. This number might seem high, but it is the result of combining all power supplies with all possible EURO regulations. The  $CO_2$  equivalent can be determined for each of the 63 modes of travel by using a life-cycle approach and by processing data in the GaBi software, according to the data sets and the EF 3.0 impact category “Global Warming” (Curran 2012, JRC-IES 2010, IKP and PE 2002, EC 2018).

## 2.2. Data preparation

We obtain 20 variables of interest from the data set used in Cappelletti et al. (2021), as described in Table 1 with 2998 resulting rows. Many of these variables are categorical, and the rest are numeric.

Table 1  
Description of the variables in the dataset

Variable	Description
Sex	Sex of respondents
Age Role	Age of respondents Role within the university Respondent department
Department Distance Min	Km covered by respondents (round trip) Time taken to reach the department
Day hot Day cold Means cold Means hot Car power	Number of days (hot season) that the respondent goes to university Number of days (cold season) that the respondent goes to university.
Car registration Passengers Locations	Means of transport (cold season) used to get to the university Means of transport (hot season) used to get to the university
First choice rent Second choice rent Third choice rent Fourth choice rent Lunch	Power supply of the car (if it has been chosen as a means of transport) Year of registration of the car
Type lunch	Number of passengers carried during the car journey
	Moving to other university locations (other than your own department) First choice alternative solutions for clean mobility
	Second choice alternative solutions for clean mobility Third choice alternative solutions for clean mobility Fourth choice alternative solutions for clean mobility
	Number of times the respondent has lunch at the university
	Type of lunch eaten at the university (brought from home or purchased)

Our goal was to understand which variables to use (reducing the number) so that the UniFG can improve its mobility services while reduce emissions.

The first step of the analysis was to transform the categorical variables into numeric variables, as obtained through LabelEncoder in sklearn (through Python). Any type of analysis requires standardized data (which we made all numerical). We used StandardScaler to eliminate the “force” linked to the scale, which is always present in sklearn, since we did not need to maintain the ordering between the points (the data were not ordinal). Analyses were conducted regarding the quantity of atmospheric emissions by the entire UniFG academic community. For this reason, trip duration from one’s home to the department (**Min**) was eliminated a priori, as were the second, third, and fourth choices of alternative sustainable mobility solutions (**Second-, Third-, and Fourth-choice rent**) since we assumed that the first choice was the most representative. At this point, the data set comprised 16 features describing transportation by the academic community.

We first tried to reduce the dimensionality of the data set through factor analysis, considering the 16 features. This analysis also allowed us to highlight the links between the different features (described through the loadings). This factor analysis was carried out with IBM SPSS using principal component analysis. We obtained eight factors, which explained 79.14% of the total variance (with eigenvalues > 0.8).

Table 2  
Component matrix of factor analysis

	<b>Component</b>							
<b>Variable</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>
<b>Sex</b>	-0.182	.241	.164	.160	-.565	.061	-.360	-.261
<b>Age</b>	-.319	.736	.111	-.302	.234	.182	.055	-.054
<b>Role</b>	.271	-.743	-.296	.246	-.214	-.109	-.105	.105
<b>Department</b>	.105	-.237	-.227	-.036	.558	.125	-.325	.423
<b>Distance</b>	.184	.402	-.244	-.153	.210	-.307	-.042	-.175
<b>Day hot</b>	-.038	-.112	.817	.075	.215	.151	-.192	-.032
<b>Day cold</b>	-.014	-.276	.807	.164	.145	.078	-.118	-.040
<b>Means cold</b>	.615	.486	.017	.555	.077	-.092	-.015	.092
<b>Means hot</b>	.619	.442	.029	.577	.079	-.092	.000	.119
<b>Car power</b>	.866	-.001	.098	-.319	-.108	.146	.007	-.030
<b>Car registration</b>	.872	-.010	.087	-.320	-.096	.159	.032	-.051
<b>Passengers</b>	.904	-.070	.078	-.268	-.087	.128	.021	-.003
<b>Locations</b>	-.016	.084	.346	-.116	-.342	-.476	.066	.349
<b>First choice</b>	-.208	.095	.059	.162	-.230	.506	.551	.401
<b>Lunch</b>	.062	-.141	.364	-.156	.176	-.522	.425	.049
<b>Type lunch</b>	.145	-.327	-.060	.308	.246	.073	.410	-.542

As we can see in Table 2 (which highlights the component matrix), the variables with the highest loadings (> 0.5) were related to the means of transport, the characteristics of these means, and the number of passengers, especially for the first factor loadings. For the second factor, the variables with highest loadings were related to the respondents' characteristics. In this way, we replaced the features with a reduced number of factors that describe the entire data set, while accepting a certain level of information

loss. However, the factors resulting from factor analysis are difficult to interpret, especially for possible analyses directly linked to mobility systems.

The main goals for the university are to understand mobility habits and to improve, where possible, by moving toward sustainable means of transport. However, given the data set's composition and the presence of variables of different types, we are able to carry out further analyzes manually selecting some variables of interest through ML techniques.

## 2.3. Analysis through Machine Learning

The data set includes variables that answer different questions about the type of respondent (such as the respondent's role and department), his or her transport habits, and his or her eating habits (lunch).

Figure 1 shows, through a heat map, the correlations among the different variables as a whole. Many of these correlations are representative of the results obtained in the factorial analysis, such as the strong link between the variables that explain the means of transport during the two seasons. Based on the different correlations, we can select blocks of variables to be used in ML algorithms to make predictions on the mobility of UniFG members.

Classification was the reference task for the analyses. Because these are multi-class classifications, we used *logistic regression* with the one-vs-rest (OvR) training algorithm in *Python*. In all cases, the training and test sets were divided into 70% and 30% portions, and a *Stratified 10-fold Cross-Validation* was performed to avoid over-fitting problems. The predictive analyses were as follows:

- Prediction of the type of car fuel (the variable [Car power]) used for moving (specifically the different types of petrol or diesel fuel) based on the variables [Age, Sex, Department, Role, Distance, Day cold, Means cold, Locations, Passengers]. The accuracy, after Cross Validation, was 79.10% with a standard deviation of 1.5%;
- Prediction of the car registration period (variable [Car registration]) based on the variables [Age, Sex, Role, Department, Distance, Day cold, Means cold, Locations, *Car power* (previously determined)]. The accuracy, after Cross Validation, was 75.70% with a standard deviation of 0.7%; and
- Prediction of the means of transport used in the hot season (variable [Means hot]) based on the variables [Age, Sex, Role, Department, Distance, Day cold, Means cold, *Car registration*, *Car power* (previously determined)]. The accuracy, after Cross Validation, was 91.30% with a standard deviation of 1.7%.

## 3. Results

We were able to combine the previous elements to try to improve sustainable mobility at UniFG. Using the previous results, we developed a system through which the university can determine the amount of  $CO_2$  equivalent for a certain semester when registering new subjects and for those already enrolled, based on small amounts of data entered in the registration system, as follows.

Information on age, gender, department, role, and residence for each subject can be acquired through the UniFG member management system (ESSE3 platform). However, the city of residence is not a functional variable since a person residing in a certain city could decide to move to Foggia. One solution could be to periodically update the ESSE3 section relating to the domicile, in order to track how the issue quantity can vary by subject over a certain time period. In this way, because we are aware of the department to which each subject belongs and his or her domicile, the distance between these two points represents the distance in km that the subject will travel with a certain vehicle. With regard to the means of transport used, if the person uses a means other than a car, the subject has chosen (whether for reasons of necessity or as his or her own choice) a sustainable mobility system. On the other hand, if the subject uses a car, then we must calculate the expected emissions.

To obtain information on the use of the car, it is possible to enter a very short questionnaire to ESSE3 with a single question, such as, "Do you use a car to go to the university? Yes/No." This question may be mandatory for new subjects wishing to enrol at the university but voluntary for those already enrolled, thus minimizing requests to guarantee a large number of answers. After obtaining information on car use, we could use the ML algorithms defined above to predict the vehicle's type of power supply and year of registration with good accuracy. In this way, we could determine the  $CO_2$ -equivalent emissions produced by subjects who use a car simply as the product of the distance travelled and the emission value returned by the GaBi for each of the EURO classes, to which we would include the subjects according to the vehicle's year of registration and type of power supply. The only variable on which we must hypothesize (always with a view to avoid burdening the questionnaire to be submitted through ESSE3) is the number of days when the subject goes to the university. However, on the basis of the training data set, we can assume that subjects who live in Foggia tend to go to the university four times a week, on average, while for those not who do not in Foggia, the number of trips tends to decrease as the distance increases (we can assume that for a distance of up to 50 km, the number of trips is three times a week, versus twice a week for a distance over 50 km). We can determine the trips per semester as the product of the weekly trips and the number of weeks in a semester (known a priori) and multiply this time value by the previous equivalent  $CO_2$  value, to obtain the emissions expected in a semester.

Through this calculation, the university could exploit the ESSE3 platform to sensitize those belonging to the academic community to the use of alternative means to cars, where possible. In particular, after setting an upper bound of emissions, we could display a sticker upon access to ESSE3 that expresses the user's expected emissions value for the following semester, with a green sticker if the emissions are below the upper bound (as shown in Figure 2), yellow if they are higher than the upper bound but within a certain deviation from this limit, and red if the expected emissions are much higher. This calculation of prospective emissions, albeit probabilistic, could allow the UniFG to sensitize its members to replacing their cars with more sustainable transportation means. Furthermore, through the latest ML model, it also would be possible to predict the transportation means used in the following semester (the summer semester, if the first use occurs in the winter one), so that these probabilistic forecasts can be improved by asking users seasonally whether they still use a car.

## 4. Conclusions

In this paper, we have developed a system that will allow the UniFG to inform its members of the level of pollutants they will emit into the atmosphere, based on their transportation habits. After training a ML system through a single question to be submitted to the academic staff, it will be possible to inform each member of how his or her means of transport affects the environment. In this way, it will be possible to sensitize the members of the academic community to the use of sustainable means of transport and to direct the rectorate's choices toward new types of incentives. Such a mechanism is not limited to the UniFG, it could be extended to any type of public or private organization. It would be sufficient for the organization's management to submit a few simple questions to its members (in order to guarantee the greatest number of answers) in order to determine the respondents' transportation habits and emissions levels. With this knowledge, it would be possible to better regulate the incentive policies for using certain types of sustainable means of transportation.

## Declarations

## Acknowledgments:

We want to thank the Rector, Pierpaolo Limone and the Vice Rector, Agostino Sevi of the University of Foggia for their concrete support and encouragement to carry out this research. We also thank the student associations for disseminating the questionnaire and encouraging students to complete it.

## References

1. Asensio OI, Alvarez K, Dror A, Wenzel E, Hollauer C, Ha S, 2020 *Real-time data from mobile platforms to evaluate sustainable transportation infrastructure. Nature Sustainability* 3:463–471.
2. Banister D, 2008 *The sustainable mobility paradigm. Transport Policy* 15(2):73–80.
3. Basu R, Ferreira J, 2020 *Understanding household vehicle ownership in Singapore through a comparison of econometric and machine learning models. Transportation Research Procedia* 48:1674–1693.
4. Becker H, Ciari F, Axhausen KW, 2017 *Comparing car-sharing schemes in Switzerland: User groups and usage patterns. Transportation Research Part A: Policy and Practice* 97:17–29.
5. Cappelletti GM, Grilli L, Russo C, Santoro D, 2021 *Sustainable mobility in Universities: The case of the University of Foggia (italy). Environments* 8(6)(57).
6. Chakhtoura C, Pojani D, 2016 *Indicator-based evaluation of sustainable transport plans: A framework for paris and other large cities. Transport Policy* 50:15–28.
7. Curran MA, 2012 *Life Cycle Assessment Handbook: A Guide for Environmentally Sustainable Products* (Hoboken, NJ, USA, John Wiley & Sons, Inc.).
8. de la Torre R, Corlu CG, Faulin J, Onggo BS, Juan AA, 2021 *Simulation, optimization, and machine learning in sustainable transportation systems: Models and applications. Sustainability* 13.

9. EC, 2018 *European Commission, Environmental Footprint Guidance document* (Guidance for the development of Product Environmental Footprint Category Rules (PEFCRs), version 6.3).
10. Hasan U, Whyte A, Jassmi H, 2020 *A review of the transformation of road transport systems: Are we ready for the next step in artificially intelligent sustainable transport? Applied System Innovation* 3.
11. IKP, PE, 2002 *GaBi 4 - Software-system and databases for life cycle engineering* (Stuttgart, Echterdingen).
12. Jeon CM, Amekudzi A, 2005 *Addressing sustainability in transportation systems: Definitions, indicators, and metrics. Journal of Infrastructure Systems* 11:31–50.
13. JRC-IES, 2010 *International Reference Life Cycle Data System (ILCD) Handbook - General Guide for Life Cycle Assessment - Detailed Guidance* (Publications Office of the European Union. Luxembourg).
14. Liang L, Xu M, Grant-Muller S, Mussone L, 2019 *Household travel mode choice estimation with large-scale data – an empirical analysis based on mobility data in milan. International Journal of Sustainable Transportation*.
15. Lopez-Arboleda E, Sarmiento AT, Cardenas LM, 2020 *Systemic approach for integration of sustainability in evaluation of public policies for adoption of electric vehicles. Systemic Practice and Action Research* .
16. Migliore M, Burgio AL, di Giovanna M, 2014 *Parking pricing for a sustainable transport system. Transportation Research Procedia* 3:403–412.
17. Nandal M, Mor N, Sood H, 2020 *An overview of use of artificial neural network in sustainable transport system. Computational Methods and Data Engineering* 1227:83–91.
18. Simons D, Clarys P, Bourdeaudhuij ID, de Geus B, Vandelanotte C, 2014 *Why do young adults choose different transport modes? Transport Policy* 36:151–159.
19. Suchanek M, Szmelter-Jarosz A, 2019 *Environmental aspects of generation y’s sustainable mobility. Sustainability* 11.
20. Tafidis P, Sdoukopoulos A, Pitsiava-Latinopoulou M, 2017 *Sustainable urban mobility indicators: Policy versus practice in the case of greek cities. Transportation Research Procedia* 24:304–312.
21. Tang T, Liu R, Choudhury C, 2020 *Incorporating weather conditions and travel history in estimating the alighting bus stops from smart card data. Sustainable Cities and Society* 53.
22. Weichenthal S, Farrell W, Goldberg M, Joseph L, Hatzopoulou M, 2014 *Characterizing the impact of traffic and the built environment on near-road ultrafine particle and black carbon concentrations. Environmental Research* 132:305–310.
23. Yang Y, Heppenstall A, Turner A, Comber A, 2020 *Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems. Computers, Environment and Urban Systems* 83.
24. Zhou X, Wang M, Li D, 2019 *Bike-sharing or taxi? modeling the choices of travel mode in chicago using machine learning. Journal of Transport Geography* 79.

# Figures

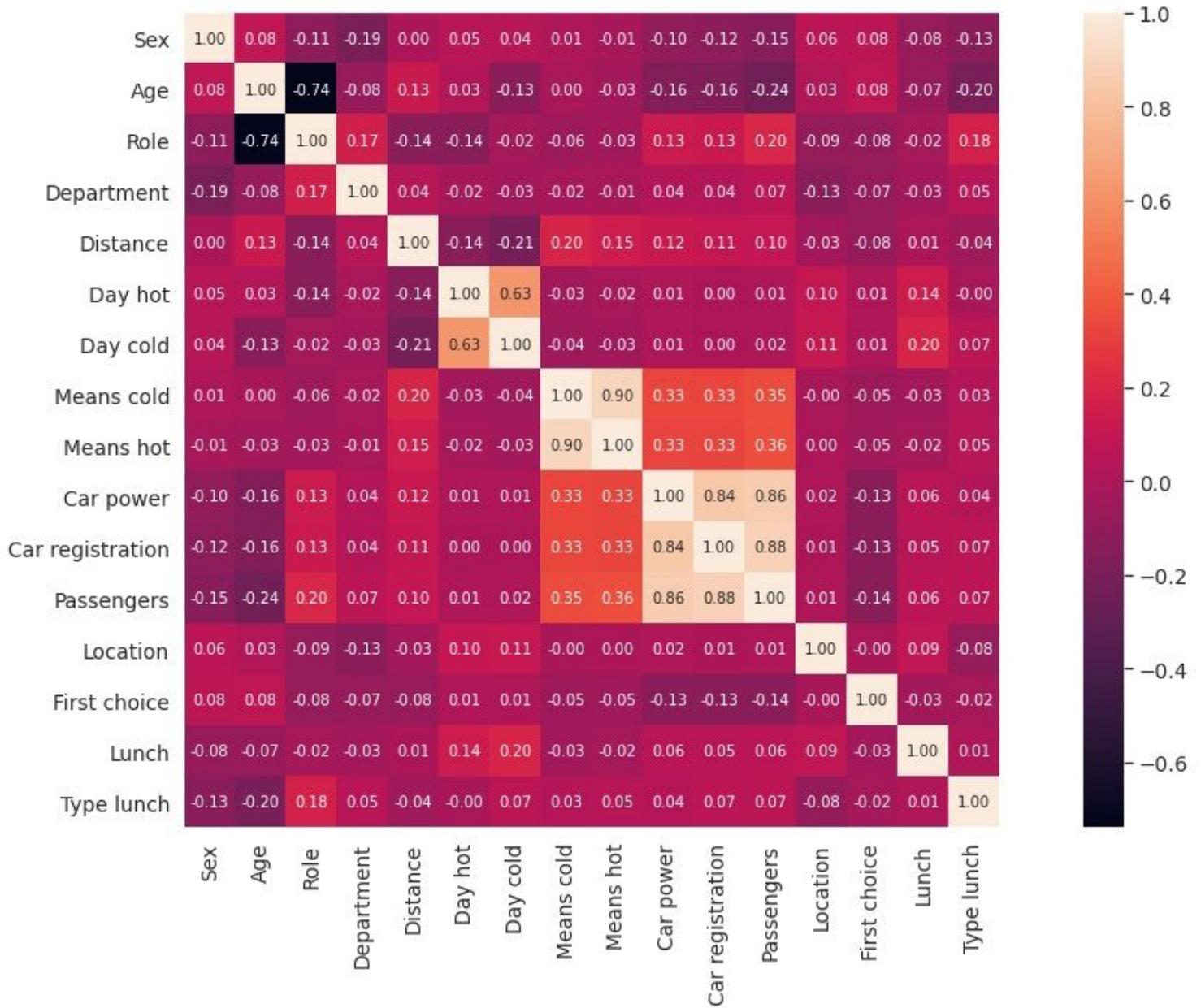


Figure 1

Heat map of correlations



KIDN

**Registrato**

PIPO PALLO



Area Riservata  
Logout  
Cambia Password

Home

Segreteria

- Test di Ammissione
- Test di Valutazione
- Esami di Stato
- Immatricolazione
- Titoli di studio
- Autocertificazione
- Pagamenti
- Dichiarazione Invalidità
- Documenti di identità

**Immatricolazione**

In questa pagina sono presentati i dati inseriti durante la procedura di immatricolazione. Se tutti i dati sono corretti, si può procedere alla visualizzazione e stampa del documento di immatricolazione.

**Dati Immatricolazione**

Anno Accademico	2017
Facoltà	Dipartimento di Studi Umanistici, Lettere, Beni Culturali, Scienze della Formazione
Tipo di Corso di Studio	Corso di Laurea
Corso di Studio	SCIENZE DELL'EDUCAZIONE E DELLA FORMAZIONE

**Dati Immatricolazione al Sistema Universitario**

A.A. Immatricolazione SU	2017/2018
Data Immatricolazione SU	03/07/2017
Ategeo Immatricolazione SU	Università degli Studi di FOGGIA - TEST

[Procedi con una Nuova Immatricolazione](#)

[Stampa Domanda di Immatricolazione](#)

Le tue emissioni sono basse



Figure 2

Example of display of the equivalent CO2 level via the ESSE3 platform