

# **Python Module - Project Description**

## **Introduction to programming for non-informatics students**

Department of Information Theory and Communication Systems

February 12, 2023

### **Contents**

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>Project Proposal I - Analysing Uber Trips</b>      | <b>2</b> |
| <b>2</b> | <b>Project Proposal II - Discovery of Handwashing</b> | <b>2</b> |
| <b>3</b> | <b>Project Proposal - Analysing Facebook posts</b>    | <b>3</b> |
| <b>4</b> | <b>Project Proposal - Geomapping</b>                  | <b>3</b> |
| <b>5</b> | <b>Multipath Fading Simulation</b>                    | <b>5</b> |
| <b>6</b> | <b>Visualisierung Hochschulstatistik</b>              | <b>7</b> |

## 1 Project Proposal I - Analysing Uber Trips

Uber has become one of the most famous unicorn companies, offering its services to more than 80 countries worldwide. To serve their customers better and provide uninterrupted services, they want to analyse the data (Uber rides) collected over time to detect patterns, such as the busiest day or the time with the lowest number of rides.

This project offers the opportunity to get familiar with the common steps of analyzing a dataset and how to draw insights from data through visualizing the relationship between different variables, solidifying the foundational rules of coding in Python.

### What will You Learn from This Python Project?

- Read a dataset and display records from it
- Use your detective's hat and uncover hidden patterns in data
- Find the relationship between different variables
- Draw insights by visualizing these relationships

## 2 Project Proposal II - Discovery of Handwashing

Prior to the nineteenth century, hygiene was frowned upon and people did not understand the risk of not washing their hands. Such a behavior caused many deaths, especially in hospitals where most doctors—unlike today—never washed their hands before surgeries.

To resolve this issue, in the mid-1800s, the Hungarian physician, Dr. Ignaz Semmelweis, started investigating the reasons behind the deaths in Vienna general hospital and those of women giving birth in particular. After analyzing data from Clinic 1 and Clinic 2, it turned out the proportion of deaths in Clinic 1 is much higher than that in Clinic 2.

By investigating, Dr. Semmelweis concluded that many of the medical students who worked at Clinic 1 also spend a lot of time in the autopsy room as a part of their studies. Thus, he realized that dealing with corpses spread harmful bacteria that was transferred to the women giving birth, thus infecting them with the deadly childbed fever - the main reason for the high mortality rates.

Analysing the available dataset, you will recreate the steps that someone did 200 years ago but this time using advanced technology.

### What will You Learn from This Python Project?

- Read a dataset and display records from it
- Display the relationship between variables through different periods
- Perform some calculations to create new variables
- Visualize the effect of handwashing on decreasing the number of deaths

### 3 Project Proposal - Analysing Facebook posts

Social media plays an important role in people's everyday life. The social media data has many uses, both in the business and social realm. Analysis of social media data would help us understand the crowd psychology and improve our business models.

We are given a piece of data on Facebook's popular posts. Please use Pandas to read the data set and perform the following analysis:

1. Find the mean and standard deviation of the number of likes for each type of post.
2. Find which time of day is most popular for posting based on the current data-set.
3. Compute the difference of likes between paid and unpaid posts.
4. Compute the correlation between "Lifetime Post Total Reach" and "Lifetime Post Total Impression".

### 4 Project Proposal - Geomapping



Geomapping involves using geographic data to create maps that illustrate how features or phenomena are related spatially. There are a number of reasons why this skill is important. One advantage of geomapping is that it can help people better understand

and analyze spatial patterns. For example, geomapping can be used to show how the distribution of certain resources, such as water or vegetation, varies across a region. This can help people make more informed decisions about how to manage those resources.

As part of the exercise, students will learn how to create simple maps by working with data sets. For example, if you want to show how the distribution of universities in Lower Saxony looks like, this can be shown with conventional methods of data analysis, but it hides certain data that only become visible through the spatial representation. In the exercise, the universities of Lower Saxony are plotted and the students are instructed to write a short text about how they interpret data or what they see. Thereby it becomes obvious that most of the universities are located in Hannover or that there are preferred locations for universities that are concentrated in the eastern part of Lower Saxony. All in all, the ability of geomapping should help to look at data(-sets) differently.

### **What will You Learn from This Python Project?**

- Create maps in Python
- Visualization of coordinates with different kinds of marker (dot, circle, rectangle)
- Customization of markers (color, symbol)
- Add popup information to markers

## 5 Multipath Fading Simulation

### 5.1 Szenario

Wir gehen von einem Übertragungsszenario mit einem Sender und einem Empfänger aus. Beide befinden sich in einem (sehr langen) Tunnel mit Breite  $B$  und Höhe  $H$ . Dieses Szenario ist in Abbildung 1 dargestellt. Der Sender befindet sich an den Koordinaten  $(x_t, 0, h_t)$  und der Empfänger bei  $(x_r, y_r, h_r)$ .

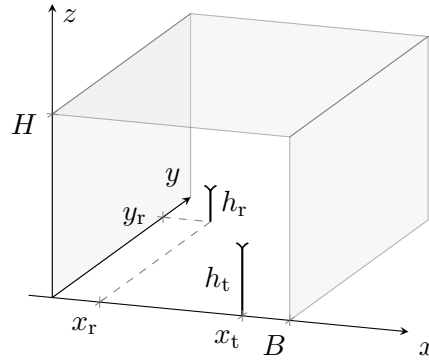


Abbildung 1: Betrachtetes Übertragungsszenario eines Tunnels mit Breite  $B$  und Höhe  $H$

### 5.2 Ziel der Aufgabe

Es soll die Leistung des empfangenen Signals am Empfänger berechnet werden. Diese soll über einen variierenden Abstand des Empfängers in  $y$ -Richtung geplottet werden.

Aufgrund der geometrischen Anordnung von Sender und Empfänger wird das am Sender abgestrahlte Signal an den Wänden (inkl. Boden) reflektiert. Diese Reflexionen überlagern sich am Empfänger und bilden das Gesamtsignal.

### 5.3 Mathematische/Nachrichtentechnische Grundlagen

Es wird das folgende (vereinfachte) Modell der Empfangsleistung  $P_r$  verwendet

$$P_r = P_t \left( \frac{c}{2\omega} \right)^2 \left| \sum_{i=1}^N \frac{A_i}{\ell_i} \exp \left( -j \frac{\omega}{c} \ell_i \right) \right|^2, \quad (1)$$

wobei die folgenden Variablen verwendet werden:

- $P_t$  Sendeleistung
- $c$  Lichtgeschwindigkeit
- $2\pi f = \omega$  (Kreis-)Frequenz
- $\ell_i$  (Gesamt-)Pfadlänge des  $i$ -ten Pfades
- $A_i$  Reflektionskoeffizient ( $A_i = 1$  für Sichtverbindung und  $A_i = -1$  für reflektierte Signale)

Für alle Betrachtungen in diesem Praktikum sollen nur die einfachen Reflexionen betrachtet werden.

## 5.4 Aufgaben

Die folgenden Aufgaben sollen im Rahmen des Praktikums in der Programmiersprache Python implementiert werden.

1. Berechnung der absoluten Pfadlängen  $\ell_i$  für gegebene Positionen  $(x_t, 0, h_t)$  und  $(x_r, y_r, h_r)$
2. Berechnung der Empfangsleistung nach (1) für gegebene Positionen von Sender und Empfänger und Frequenz  $f$
3. Plot der Empfangsleistung über  $y_r$  und Vergleich mit der Empfangsleistung ohne Reflexionen

## 6 Visualisierung Hochschulstatistik

### 6.1 Szenario und Ziel der Aufgabe

Das Statistische Bundesamt veröffentlicht regelmäßig verschiedene Daten und Statistiken zu hochschulbezogenen Themen in Deutschland. Diese können zum Großteil frei über das GENESIS-Portal [1] online abgerufen werden.

Im Rahmen dieses Projekts sollen ausgewählte Daten angemessen grafisch dargestellt und verglichen werden. Insbesondere betrifft dies die Entwicklung der Studierendenzahlen (21311), des wissenschaftlichen Hochschulpersonals (21341) und die finanziellen Ausgaben der Hochschulen (21371). Die angegebenen Zahlen in Klammern entsprechen dem Code des jeweiligen Datensatzes in der GENESIS-Datenbank.

Für den Vergleich verschiedener Studiengänge/Fachbereiche wurden exemplarisch Elektrotechnik und Betriebswirtschaftslehre (BWL)/Wirtschaftswissenschaften ausgewählt.

### 6.2 Aufgaben

Die folgenden Aufgaben sollen im Rahmen des Praktikums in der Programmiersprache Python implementiert werden. Für die grafischen Darstellungen soll eine geeignete Form der Darstellung gewählt werden. Es gibt keinerlei Einschränkungen bei der genauen Art der Darstellung, solange die Sachverhalte und Zusammenhänge gut erkennbar sind.

Die benötigten Daten werden bereits vorsortiert und bereinigt in Form von CSV-Dateien bereitgestellt.

1. Einlesen der Daten aus den bereitgestellten CSV-Dateien
2. Geeignete grafische Darstellung der zeitlichen Entwicklung der Studierendenzahlen und Frauenquote in den verschiedenen Studiengängen
3. Geeignete grafische Darstellung der zeitlichen Entwicklung der absoluten Zahlen und der Frauenquote des wissenschaftlichen Hochschulpersonals für die verschiedenen Fachbereiche
4. Geeignete grafische Darstellung der zeitlichen Entwicklung der Hochschulausgaben der einzelnen Bundesländer (sowohl die Gesamtausgaben als auch die Ausgaben der einzelnen Fachbereiche sowie deren Anteil an den Gesamtausgaben)
5. *Bonus:* Gegenüberstellung der Entwicklungen von Studierendenzahlen, Personalzahlen und Ausgaben

## Literatur

- [1] “GENESIS-Online, Die Datenbank des Statistischen Bundesamtes”, Statistisches Bundesamt (Destatis). (2021), Adresse: <https://www-genesis.destatis.de/genesis/online> (besucht am 18.08.2022).