

Fahrradverkehr in München

Prüfungsleistung

Grundlagen Data Science

des Studienganges Informatik

an der Dualen Hochschule Baden-Württemberg Stuttgart

von

Felix von dem Berge, Felix Rakoczy und Lars Huzel

09.01.2025

Bearbeitungszeitraum

7 Wochen

Matrikelnummern, Kurs

8927065, 1764797, 1628025, INF22D

Gutachter der Dualen Hochschule

Prof. Dr. Martin Lachmair

Aufgabe b

1. Wie hat sich das Fahrradaufkommen an den verschiedenen Zählstellen in München seit 2008 entwickelt?

Ziel: Trends im Fahrradverkehr analysieren, um mögliche Veränderungen im Mobilitätsverhalten der Münchner Bevölkerung zu erkennen.

2. Gibt es saisonale Muster im Fahrradverkehr?

Ziel: Untersuchung, wie stark sich das Fahrradaufkommen je nach Jahreszeit verändert, um saisonale Schwankungen zu identifizieren und besser auf die Bedürfnisse von Radfahrern einzugehen.

3. Welche Zählstellen zeigen das größte Wachstum im Fahrradaufkommen, und welche Faktoren könnten dies beeinflussen (z. B. Infrastrukturverbesserungen, Bau neuer Radwege)?

Ziel: Verstehen, welche Maßnahmen das Fahrradfahren fördern und wie die Infrastruktur gezielt verbessert werden kann.

Angabe der Verwendung von KI

Aufgabe a

Wir haben zunächst überlegt, was alles in eine deskriptive Statistik gehört und was beim Bereinigen der Daten alles zu beachten ist. Um dafür eine Auflistung zu erhalten haben wir ChatGPT folgendes gefragt:

Erstelle eine sinnvolle deskriptive Statistik der Daten. Berechne dazu die Daten entsprechend.

Anschließend wollten wir wissen, wie die Daten zunächst im Detail bereinigt werden können. Dazu haben wir folgendes gefragt:

Wie kann man in Python in einer CSV-Datei nach Duplikaten suchen?

Wie kann man überprüfen, ob die Datenformate korrekt sind?

Wie kann ich Zeilen herausfiltern, die in einer bestimmten Spalte NA als Wert besitzen?

Wie kann man die Zeilen herausfiltern, bei denen das uhrzeit_start und das uhrzeit_ende gleich ist?

Wie kann man die Zeilen herausfiltern, bei denen richtung_1 + richtung_2 nicht gesamt ergibt?

Nachdem alle Daten bereinigt waren, wollten wir nun zu deskriptiven Statistik übergehen, weshalb wir folgendes gefragt haben:

Erstelle mir eine deskriptive Statistik für die bereinigten Daten.

Aufgabe b

Keine Verwendung von KI

Aufgabe c

Für diese Aufgabe konnte bereits Teile des Codes aus Aufgabe a) wiederverwendet werden, z.B. wie die Daten eingelesen werden können oder wie nach bestimmten Zeilen gefiltert werden kann.

Daher beziehen sich die Anfragen in dieser Aufgabe nur darauf, wie man mit Matplotlib auf Basis der bereinigten Daten die entsprechende visuelle Auswertung erstellen kann. Mit Hilfe dieser Fragen konnte ein Skript erstellt werden, das dann PNGs für die gesamte Analyse erzeugen konnte.

Zunächst wollten wir wissen, wie generell ein Skript in Python aussehen kann, um solche visuellen Analysen zu erstellen. Dazu haben wir den entsprechenden Code aus Aufgabe a) recycelt und Code-Schnipsel aus den folgenden Abfragen verwendet.

Wie kann man mit Matplotlib in Python visuelle Auswertungen erstellen, die im PNG-Format gespeichert werden können?

Um die PNG-Bilder ansprechender zu gestalten, haben wir anschließend von ChatGPT einige Stylings für Matplotlib erstellen lassen.

Wie kann man in den Visuellen Auswertungen die X- und Y-Achse korrekt beschriften und einen entsprechenden Titel eingeben?

Wie kann in der visuellen Auswertung auf den Balken die genaue Anzahl angezeigt werden, die die Werte der Y-Achse widerspiegelt?

Wie kann ich diese Zahl für jeden Balken immer auf der gleichen Höhe anzeigen?

Wie kann für die visuellen Auswertung ein Titel angegeben werden?

Bei der Frage nach den saisonalen Mustern wollten wir die Quartale verwenden, die in etwa mit den Jahreszeiten gleichgesetzt werden können.

Wie kann ich mein aktuelles Skript anpassen, um das Datum in die Quartale Q1 bis Q4 einteilen zu können?

Aufgabe d

Um übersichtlichere Daten für die Präsentation zur Verfügung zu haben, wurden die Skripte von exec_3-2 und exec_3-3 so erweitert, dass sie eine Übersichtsgrafik des Radverkehrsaufkommens pro Quartal (Jahresvergleich), das gesamte Radverkehrsaufkommen aller Zählstellen mit Wachstumsrate, eine Gesamtübersicht der

Zählstellen und die einzelnen Zählstellen mit Wachstumsrate darstellen. Dabei wurde jeweils das gesamte Programm in GPT hochgeladen und mit den Prompts verändert.

exec_3-2:

Dateiupload

Kannst du hier noch eine Gesamtübersicht über die Quartale als Grafik erstellen?

exec_3-3:

Dateiupload

Kannst du das so erweitern, dass noch eine Grafik erstellt wird, in der eine Gesamtzahl für alle Standorte erstellt wird?

Erweitere diese Grafik um die durchschnittliche Wachstumsrate und gebe zusätzlich den Durchschnitt der letzten 8 Jahre an.

Bitte erstelle eine Grafik, in der alle Jahre kumuliert für jede Zählstelle dargestellt wird.

Erweitere die Grafiken der einzelnen Zählstellen um die Wachstumsrate.

Bitte entferne alle Punkte der Wachstumsrate, die über 160% und unter -50% liegen.

Aufgrund einzelner Ausfälle von Messstationen haben sich Wachstumsraten von bis zu 800% eingeschlichen, die nicht repräsentativ sind und daher zu verwerfen sind.

Aufgabe e

Um zunächst einen Boilerplate-Code für die Dash App zu erhalten, haben wir zunächst ChatGPT die Aufgabe gestellt. Diesen Code konnten wir dann weiterentwickeln und auf unsere Bedürfnisse anpassen:

Erstelle eine Dash-App, die den jährlichen Verlauf der Zählungen pro Jahr darstellen kann.

Da im Code nur ein einziges Dropdown vorhanden war, wollten wir wissen, wie man mehrere Dropdowns in unseren Code einbinden können:

Ich möchte mehrere Dropdowns integrieren. Ich möchte über die Dropdowns Jahr, Zählstelle und Richtung auswählen können. Wie würde der Code aussehen?

Da die Zählstellen nicht für jedes Jahr gleich sind, wollten wir, dass sich die Werte im Zählstellen-Dropdown aktualisieren, falls ein Jahr mit anderen Zählstellen ausgewählt wird:

Ich möchte zudem, dass sich die Zählstellen im Dropdown ändern, sobald ein anderes Jahr im Jahr-Dropdown ausgewählt wird.

Da momentan die Werte von allen Zählstellen in einem Graph dargestellt werden, wollten wir nun die Daten abhängig von der Zählstelle zu filtern. Dafür haben wir folgendes gefragt:

Ich möchte alle Daten von einem Jahr der Zählstelle ‚Arnulf‘. Gib mir dafür bitte die Query.

Am Ende wollten wir die Webseite noch ein wenig aufwerten, weshalb wir noch ein paar Designfragen gestellt haben:

Wie kann man den Font-Style ändern?

Wie kann man ein html.Label Element stylen?