

Kurzbericht Lukas Trölenberg

Beschreibung der eigenen Rolle und Aufgabenbereiche

In Rahmen unserer Projektarbeit haben wir uns gemeinsam einen Datensatz auf Kaggle ausgesucht. Anschließend habe ich zusammen mit Tom eine zentral organisatorische Aufgabe übernommen: das Erstellen einer To-Do-Liste mit Aufgabenzuteilung. Im Anschluss habe ich mich potenziellen Verzerrungen des Datensatzes und der Analyse Fehlender Werte und Ausreißern gewidmet. Da die anfänglichen Analysen keine signifikanten Auffälligkeiten im Datensatz zeigten, habe ich ergänzend potenzielle Analyseverfahren ausgearbeitet, die bei entsprechenden Ergebnissen zur Anwendung gekommen wären. Zudem habe ich eine Data Dictionary, mit einschließlich der neu implementierten Features, angelegt und mich mit der kritischen Hinterfragung von unserem Projekt auseinandergesetzt. Hinzu kommen die möglichen Schritte für eine erweiterte Analyse. Nebenbei habe ich an der PowerPoint-Präsentation mitgewirkt.

Reflexion der größten Herausforderungen

Anfänglich war bereits die Suche nach einem gut geeignetem Datensatz eine große Herausforderung. Nach drei unterschiedlichen Datensätzen haben wir uns schließlich auf einen geeinigt, der unseren Vorstellungen gerecht wurde. Viele der in Erwägung gezogenen Datensätze waren nichts aussagend und wir konnten kein greifbares Business-Problem feststellen.

Bei der Erstellung der To-Do-Liste und der Aufgabenverteilung kam unsere mangelnde Erfahrung mit Datenanalysen zum Vorschein und wir konnten kaum die Aufgaben nach Stärken und Schwächen zuordnen. Die Ausarbeitung von potenziellen Verzerrungen gestaltete sich auch durch Mangel an Erfahrung und fachlichen Kompetenzen schwierig. Eine weitere Herausforderung stellte der Umgang mit Analysen in Python dar. Sämtliche Befehle musste ich erst heraussuchen oder mir von KI geben lassen. Die Erstellung der Data Dictionary verlief reibungslos, bis auf ein paar Komplikationen mit dem Markdown-Format. Die größte Herausforderung für mich war es sich mit der kritischen Hinterfragung unserer Ergebnisse auseinanderzusetzen und sinnvolle Fragen zu formulieren, weil hier ebenfalls die Erfahrung rund um die Analyse von Daten gefehlt hat und somit die eigene Meinung auf den ersten Blick eher mager ausfällt.

Ein weiteres allgegenwärtiges Problem war die kollektive Bearbeitung des Notebooks. Durch die zahlreichen Änderungen und Aktualisierungen auf GitHub kam es vor, dass ein Notebook mehrfach hintereinander aktualisiert wurde und so viele Änderungen verloren gingen, die dann im Nachhinein wieder eingefügt werden mussten.

Analyse der eigenen Learnings

Aus dem Projekt nehme ich mit, dass bei einer solchen Gruppenarbeit besonders die Kommunikation untereinander im Mittelpunkt steht, um einen reibungslosen Ablauf zu garantieren. Zudem wurde mir klar wie Data Scientists ihren workflow strukturieren und ich

Gruppe 8: Engel, Nolepa, Schneider, Trölenberg
Kurzbericht Tom Schneider
19.11.2025

konnte etwas an praktischer Übung mit Daten sammeln. Insbesondere bleibt mir die Auslegung von Daten in Bezug auf deren Interpretation hängen, z.B. wie einfach es ist durch kleine Anpassungen die Aussagen von Daten um 180 Grad zu drehen.