Einreichung Projektvorschlag

für

Data Analytics

(17.11.25 – 12.12.25)

Projektersteller/Gruppe: Group E: East/Lynam-Smth

Lynam-Smith, Miles

Daten: 24 November 2025

Projektidee, Beschreibung:

The Himalayan database is rich in historical value, detailing the peaks, expeditions, climbing statuses, and geographic information of numerous Himalayan summits.

We will explore this data dating back to the start of the last century.

Our approach will be clean and tidy the data - using Numpy, Pandas in Jupyter notebook, before the data analysis takes place. We will record both the steps taken and the results plus any lesson learnt. The data analysis will focus on the inspiration taken from Tidy Tuesday in Github which is detailed in the following section; using MatPlotLib, Seaborn, Plotly (including Dash), and SpaCy.

Our purpose is to continuation of the work of Elizabeth Hawley, a longtime journalist based in Kathmandu, who dedicated most of her life to archiving expeditions in the Nepal Himalaya, whilst demonstrating our newly acquired data science skills.

https://www.himalayandatabase.com/downloads/Himalayan%20Database.zip

Projektschritte, Aufgaben zum Projekt (falls schon vorhanden):

Data Import & Understanding:
Load all relevant Himalayan Database tables into Pandas DataFrames and explore their structure (data dictionary, columns/fields, first rows, data types).

Exploratory Data Analysis (EDA):
Used descriptive statistics and visualizations to explore the data, detect outliers, missing values, and understood important variables (e.g., peak heights, climbing status, routes used).

Data Cleaning & Transformation:

Remove or impute missing values, standardize naming conventions/categories (e.g., names of peaks and countries), convert data types for dates/times, numbers, etc, engineer new features, e.g., create an "expedition success" variable based on all route outcomes.

Documentation & Version Control:
Record every step in Jupyter Notebook using comments and markdown. Document issues and solutions and version key milestones using GitHub.

External Data Consistency Check:
Check data reliability by cross-referencing some expeditions/peaks with external sources (publications, the Himalayan Database website).

Reflection & Lessons Learned:
At the end, reflect on what worked well in the cleaning and analysis process, where trouble arose, and what insights you gained.

Presentation/Reporting:
Prepare key visualizations, analysis results, and main messages for a presentation or poster.

What is the distribution of climbing status (PSTATUS) across different mountain ranges (HIMAL_FACTOR)?
Which mountain range (HIMAL_FACTOR) has the highest average peak height (HEIGHTM)?
What is the distribution of peak heights (HEIGHTM) for peaks that are open (OPEN) versus those that are not?
Which climbing routes (ROUTE1, ROUTE2, ROUTE3, ROUTE4) have the highest success rates (SUCCESS1, SUCCESS2, SUCCESS3, SUCCESS4) across all expeditions?
How does the use of supplemental oxygen (O2USED, O2NONE) affect summit success rates?How often does bad weather (TERMREASON = 4) play a role in termination compared to technical difficulty (TERMREASON = 10)?
Are expeditions with no hired personnel (NOHIRED) associated with higher or lower death rates?