

Data Aggregation Gipfelstürmer

Feature-Extraktion

Folgende Attribute wurden aus den Berichten extrahiert und aufbereitet:

- **name**
- **id**
- **categories:** Kategorien werden aus den Schwierigkeitsgraden abgeleitet.
- **difficulty_per_category:** Schwierigkeiten sind in unterschiedlichen Skalen angegeben und werden auf eine einheitliche Skala harmonisiert.
- **date:** Das Datum wird als Unix-Timestamp übernommen.
- **duration:** Die Dauer wird in Stunden (float) gespeichert. Ein Tag wird dabei als 24 Stunden interpretiert.
- **ascent**
- **descent**
- **peaks:** Alle Wegpunkte, welche als Icon `ico2_peak_s.png` als verwenden, werden als Gipfel übernommen.
- **views**

Auswertung

Für die Rangliste wurden nur die Touren berücksichtigt, welche folgende Kriterien erfüllen:

- Die Tour muss mind. einen Gipfel enthalten.
- Die Tour muss mind. 1000 Views aufweisen.
- Die Tour muss mind. eine Kategorie (also mind. eine Schwierigkeit) enthalten. Zudem dürfen die Kategorien "Eisklettern" und "Skifahren" nicht enthalten sein.
- Keine der angegebenen Schwierigkeiten aus den Kategorien darf höher als Stufe 5 gewertet sein.
- Der Abstieg sowie Aufstieg sollte geringer als 1800m sein.
- Die Dauer der Tour sollte unter 24 Stunden liegen.

Danach wurden die 10 meist vorkommenden Gipfel ausgewählt und nach deren Anzahl sortiert. Als zweites Sortier-Kriterium wurde die Summe der Aufrufe aus allen referenzierten Touren verwendet.

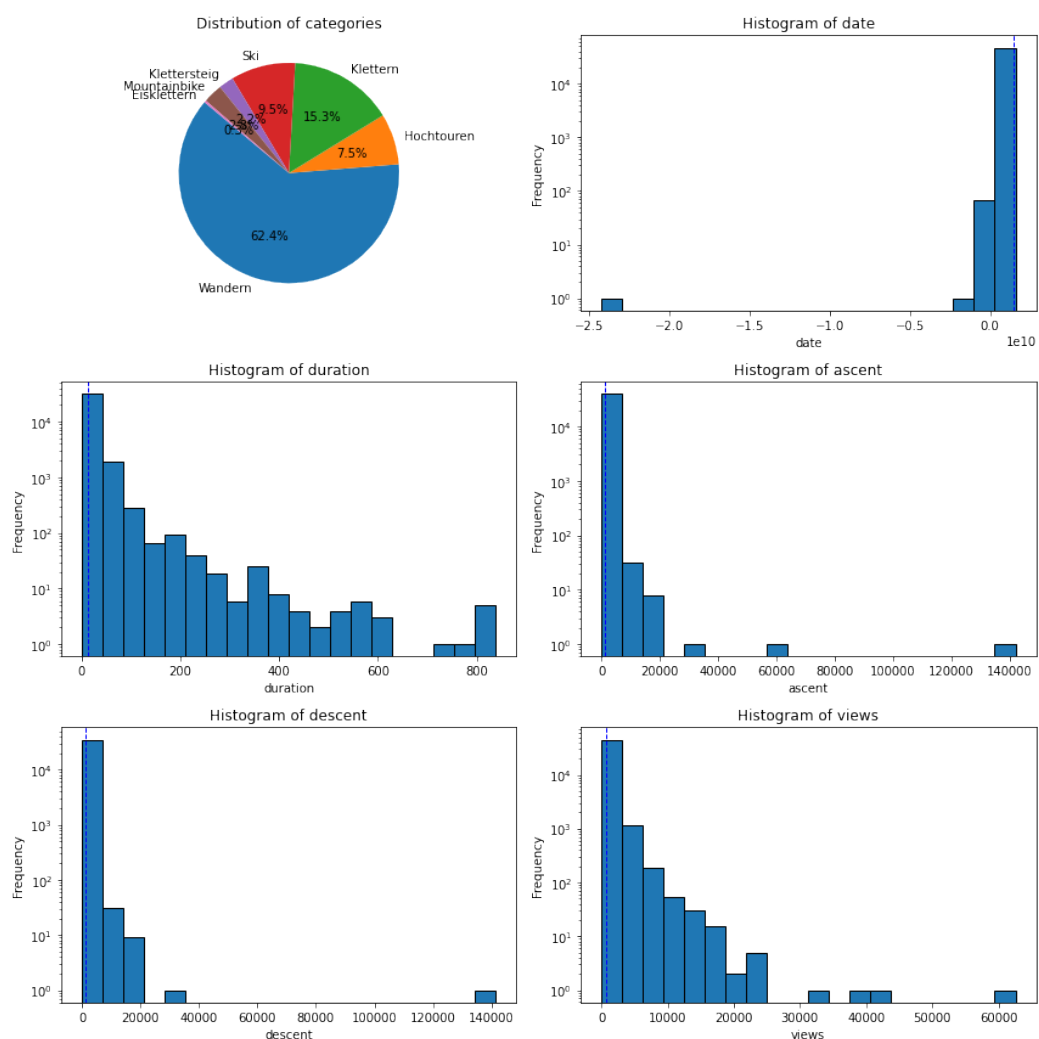
Gipfel	Anzahl Berichte	Anzahl Aufrufe (view)
Säntis	24	62595
Fronalpstock	18	35888
Speer	18	31294
Grosser Mythen	17	38075
Kleiner Mythen	16	33896
Schrattenflue - Hängst / Hengst	13	42862
Rossberg - Gnipen	13	30439
Marwees - Ostgipfel	13	23098
Tomlishorn	13	19945
Rigi Kulm	13	19053

Datenanalyse

Die Datenanalyse ergab, dass ca. 7% der Touren keine Kategorie-Zuordnung aufweist. Bezüglich der Datumsangaben fanden sich bei 83 Touren keine Einträge, wobei einige Datumsangaben unrealistisch erscheinen. Etwa ein Viertel der Touren wies keine Angaben zur Dauer auf, während rund 15% keine Aufstiegs- und 25% keine Abstiegsangaben hatten. Einzelne Touren verzeichneten keine Aufrufe. Bezüglich des Schwierigkeitsgrads hatte ein Teil der Touren keine Angaben, wobei der Schwierigkeitsgrad 2 am häufigsten war.

Generell scheint die Datenqualität gut und der Datensatz eignet sich für weitere Analysen. Der hohe Anteil fehlender Werte lässt gewisse Attribute nicht ohne weitere Massnahmen direkt nutzen. So müsste z. B. bei der Verwendung der Dauer oder des Abstiegs die fehlenden Daten zuerst noch bereinigen oder mithilfe der Tourenbeschreibung aufwändig rekonstruieren.

Detailliertere Analysen der Daten sind im Jupyter-Notebook zu finden.



Aggregation

Die Tourenberichte wurden nach Kategorie gruppiert. Anschliessend wurde für den Auf- und Abstieg sowie die Dauer jeweils der Durchschnitt berechnet:

Kategorie	Dauer	Aufstieg	Abstieg
Eisklettern	10.9	585.488	559.568
Hochtouren	19.0846	1528.671	1589.697
Klettern	11.34562	1343.077	1362.833
Klettersteig	12.53114	1182.761	1192.603
Mountainbike	9.142399	1196.653	1213.177
Ski	6.300985	1164.556	1272.905
Wandern	7.659494	1017.288	1011.449