Project objective

Investigate the correlation between Strava segment activity and weather conditions.

Context

Strava is widely used by atheletes around the world to keep track of their sports activities. This project focusses on cycling in specific. Moreover, usings so-called segments, users can compete against each other on certain specified tracks.

Strava segments may differ in popularity based on the weather. Our aim is to check whether a correlation between location activity and the type of weather exists.

Approach

First, using the Strava API, segment ranking data is used to calculate daily activity frequency per segment for a given timeframe. Next, weather data from these days and locations are obtained from the KNMI, as well as the KNMI weather warnings. This is used to calculate self-defined weather categories based on rain, wind, and the warnings.

For a given weather type, filter the days with that specific weather type. Then, filter the strava data for those days and generate a heat map displaying the activities per segment. A reference output is generated, so that further weather dependent outputs can be compared and interpreted.

Different modules will be created:

- A module to obtain data from Strava
- A module to obtain weather data
- A module to generate heat maps
- A module to filter for days with certain weather types

Requirements

- Obtain weather data and KNMI warnings
- Use Strava API to obtain segment activity data
- Function to calculate weather categories based on wind, rain, and KNMI warnings
- Strava activity is specified to cycling.
- Pipeline output should be verifiable.
- Reference output to compare and interpret weather dependent outputs.

Constraints

- Weather data is available for The Netherlands only, so Strava activities are only needed for The Netherlands.
- Not all people who do sports use Strava, and not all Strava users publish their activities, so not all activities on the segments are registered
- Need to take the growth/decline in Strava users over the time frame into account, or recognise it as a possible influencing factor.
- time frame is specified for the year 2019

Resources

- Strava api
- KNMI weather data
- Google Maps

Priorities

1. Collect and filter data both Strava and weather data
2. Data handling and preparation for verification
3. Data verification
4. Data visualization

Results expected

Segment activity is dependent on weather type

Design Objectives

- Design for maintainability
- Desgin for reusability
- Design for extensibility

Design Strategy

- Use scrum
- Use separate modules to decrease coupling
- Use Test Driven Development

Critical Features

- module that obtains Strava data
- module that obtains weather data
- module that filters all the data
- function that generates a reference output
- module that displays the data

Risk factors

- reference output is not representative
- data cannot be validated
- Unclear expectations and/or results

To-avoids

- Waterfall method, not responding to change
- not following the plan
- losing view of what other group members are working on

Design validation and evaluation

Every week, the team has a meeting to evaluate on the week before and reflect on itself on group dynamics. Possible action points are written down for the next week.

Test approach

- create unit tests using pytest
- use CI pipeline on GitLab
- possible manual testing

Test planning

Because we chose Test Driven Development, testing is done alongside code writing. Also, team members code review each other's code.

Test validation

Test outputs will be compared with expected results. Also, using the points from the data validation section below, we will try to validate the outputs.

Code base and documentation

GitLab will be used to manage code versions, store documentation, and enable a Continuous Integration pipeline.

Implementation planning

A timetable is provided in the documentation repository

Data Validation

- Heat map of segment activity for different weather types
- Filtering of obviously false data is mostly done by Strava
- If a KNMI warning is given, there should be very little activities