COM-480: Data Visualization

E. Michelin, A. Doukhan, F. Dumoncel

March 27, 2024

Milestone 1 (29th March, 5pm)

E. Michelin, A. Doukhan, F. Dumoncel

Datasets

In this project, we will utilize two datasets:

1. The first dataset is sourced from a well-known GitHub repository, owned by Jeff Sackmann, comprising several CSV files that aggregate extensive statistics on tennis players and matches from the 2000s. This dataset will be instrumental in analyzing and tracking the evolution of CO_2 emissions over the past two decades.

   These datasets primarily contain information about rankings, matches, and players over the past 20 years.

2. The second dataset was developed by our team through meticulous scraping and calculation of pertinent values, accessible here. We extracted data from the official ATP website to compile a comprehensive list of tournaments held throughout the season. This dataset will enable us to identify the most efficient travel itinerary for the season, specifically, the sequence of tournaments that results in the lowest possible CO_2 emissions.

   The ultimate goal is to construct a graph where cities hosting tournaments serve as nodes, and travel modes between these cities are represented as edges. For each travel mode, we will also calculate the associated emissions and the duration of travel. To achieve our goal, we operate under several assumptions:

   - Airplane travel is always an option since tournaments typically occur in major cities.
   - Train and car travel durations are sourced from the Google Maps API.
   - The average speed of an airplane is approximately 880 km/h.
   - An additional 3 hours are factored into airplane travel to accommodate boarding, commuting to and from the airport, and other related activities.

Problematic

The overarching theme of this project is the intersection of professional tennis and CO_2 emissions. The modern tennis tour is characterized by its global footprint, with tournaments scheduled in diverse cities worldwide on a weekly basis — or over two weeks for major tournaments. This relentless calendar necessitates that the top 100 players frequently travel vast distances to compete, incurring significant carbon emissions in the process. Our objective is to quantify, visualize, and evaluate these emissions to foster greater environmental awareness among players. By providing detailed insights

into their carbon footprint, we aim to encourage athletes to consider more sustainable approaches to planning their season, potentially leading to a reduction in overall emissions associated with professional tennis.

The primary aim of this project is to offer meaningful insights towards fostering a more eco-friendly tennis tour, not only for the players but also for fans, journalists, tournament directors, and anyone else with a keen interest in tennis. This initiative seeks to engage a broad spectrum of the tennis community in a dialogue about sustainability, encouraging all players to consider greener practices and strategies. Ultimately, the project aspires to contribute to the evolution of tennis into a more environmentally conscious and sustainable sport, benefiting everyone involved and the planet at large.

Exploratory Data Analysis

The preprocessing scripts for our project are located in the preprocessing/ directory. This section of our project repository houses the essential logic for processing the list of tennis tournaments and computing critical environmental impact metrics. Key computations include:

- **Distance Calculation:** Using the *geopy*[1] library, we calculate the geographical distance between cities hosting tennis tournaments. This step is crucial for estimating travel distances that players might cover during the tour.

- **CO2 Emissions Estimation:** Leveraging the *transport-co2*[2] library, we compute the estimated CO2 emissions in tons for different modes of transportation − airplane, car, and train − between any two given cities (city a and city b). This computation allows us to assess the environmental impact of travel within the tennis tour and explore more sustainable travel options.

For those interested in diving deeper into our data, exploratory data analysis of the dataset can be found in the notebook located at /notebooks/m1.ipynb. This analysis provides initial insights and visualizations that guide the project's focus on reducing CO2 emissions associated with the tennis tour.

Related work

The dataset we're utilizing, created by Jeff Sackmann, serves as the backbone for the Ultimate Tennis Statistics website. This platform is the go-to online resource for tennis aficionados seeking insights and statistics on virtually anything related to the sport over the past two decades. It offers a comprehensive and detailed exploration of players, matches, tournaments, and various statistical analyses that appeal to fans, analysts, and anyone with a keen interest in understanding the nuances of tennis.

To our knowledge, there isn't an existing online visualization that explicitly showcases the environmental impact of the ATP Tour, making our approach both original and pioneering. As avid tennis enthusiasts, our passion for the sport fuels our desire for its longevity and continued success. However, as conscientious global citizens, we are equally committed to the preservation of our planet. This dual commitment drives our project's core mission: to illustrate the carbon footprint associated with the global tennis circuit and to advocate for a more sustainable future for the sport.

Our goal is not only to raise awareness but also to catalyze change within the tennis community. Through our efforts, we hope to inspire players, organizers, and fans alike to embrace and implement

greener practices, demonstrating that a more eco-friendly tennis season is not only necessary for our planet but also achievable.

References

1.   GeoPy: Python geocoding toolbox. Kostya Esmukov, 2023, November 23.
2.   Transport-co2: Calculate CO2 emissions for personal mobility. Brylie Christopher Oxley, Markus Schepke, 2020, May 20.