Uber Nairobi Ambulance Perambulation Challenge

Can you use ML to create an optimised ambulance deployment strategy in Nairobi?

Competition description

Road traffic collisions are the number one killer of children and young adults ages 5-29, and 8th leading cause of death worldwide. Post-crash care is one of the five pillars of road safety and a critical component for reducing morbidity and mortality.

When it comes to emergency response to road accidents, every second counts. With heavy traffic patterns and the unique layout of the city, finding the best locations to position emergency responders throughout the day as they wait to be called is critical in a city like Nairobi.

We've collected information on thousands of traffic accidents that have occurred in Nairobi, Kenya in 2018 and 2019. For this competition, you must use the training data (recorded crashes up to June 2019) along with supplementary data from Uber Movement, road survey data and weather patterns to identify patterns of risk across the city. You must then use these findings to place six virtual ambulances around the city, moving them around throughout the day with the goal of minimising the distance travelled when responding to crashes during the test period.

About Uber Movement (movement.uber.com)

Uber Movement is a website that helps urban planners, city officials, riders and the public better understand the transportation needs of their cities. Presently, Movement shows average travel times between zones in a city, derived from anonymous and aggregated on-trip data from Uber vehicles. Uber is planning the next set of functionality and datasets that will be added to Movement in the coming months. Uber Movement is available for 22 cities around the world.

About the World Bank Development Impact Evaluation Department (https://www.worldbank.org/en/research/dime)

The World Bank's Development Impact Evaluation (DIME) group generates high-quality and operationally relevant data and research to transform development policy, help reduce extreme poverty, and secure shared prosperity. It develops customized data and evidence ecosystems to produce actionable information and recommend specific policy pathways to maximize impact. The road traffic crash data used for this challenge was produced thanks to funding provided by the UK Foreign, Commonwealth & Development Office (FCDO) through the ieConnect for Impact program.

About Flare
Flare builds 911 systems for places in the world without one. More than 5 billion people live in cities and countries where calling for an ambulance during an emergency like during a road accident is not a reality. Flare brings together available ambulances onto a centralised technology platform. It runs a 24/7 professional dispatch centre where their rescue.co members can dial a single number or press an SOS button to instantly reach help. In Kenya, before Flare launched, it used to take 162 minutes (nearly 3 hours!) to receive care after an emergency. Today, Flare's average response time is 15 minute in Nairobi and increasingly they are responding in 2, 3, or 4 minutes.

@RescuebyFlare

Social media presence

- Handles: Twitter, Facebook, LinkedIn, Instagram
- Hashtags (if applicable)

Evaluation

Scoring for this challenge will be slightly different to normal. Instead of predicting crash locations or frequency, you must instead submit a location schedule for **six** ambulances every 3 hours over the test period. For each crash recorded, we find the distance from that crash to the nearest ambulance according to your schedule. The final score is the total distance travelled (in decimal degrees) by all ambulances over the entire test period. Lower is better. An example of this scoring on a validation set is included in the starter notebook.

The submission file lets you specify the locations of each ambulance. Let's consider an example with just one ambulance:

date	A0_Latitude	A0_Longitude
2019-06-01 00:00:00	-1.3	36.7
2019-06-01 03:00:00	-1.3	36.7
2019-06-01 06:00:00	-1.3	36.7
2019-06-01 09:00:00	-1.32	36.62
2019-06-01 12:00:00	-1.32	36.62
2019-06-01 15:00:00	-1.32	36.62
2019-06-01 18:00:00	-1.3	36.7
2019-06-01 21:00:00	-1.3	36.7

If a crash occurs at (-1.4, 38.0) at 15:43 on 2019-06-01, the distance to the ambulance A0 (which is at (-1.32, 36.62) at that time) is 1.38232 (for simplicity we ignore any geographic projections, and assume 1 degree latitude = 1 degree longitude since Nairobi is so close to the equator). The final score is the sum of the distance to the closest ambulance for each crash in the test period.

Prizes

1st Place: $3,000 USD

2nd Place: $2,000 USD

3rd Place: $1,000 USD

Timeline

Competition closes on 24 January 2021.

Final submissions must be received by 11:59 PM GMT.

We reserve the right to update the contest timeline if necessary.

Data

The data consists of crashes identified by the World Bank DIME research team and by Flare. ‘Train.csv’ provides time and location for 6071 crashes in the training period (2018-01-01 to 2019-06-01).

You are asked to determine the locations for **six** different ambulances to be placed in order to minimize the distance to any reported crashes. Ambulances can be assigned a new location every three hours. Scoring is based on the distance from each crash in the test period to the nearest ambulance (see the Evaluation section for specifics).

Additional data is also provided:

- **Weather.csv** has daily weather based on the GFS dataset [link] - https://www.dropbox.com/scl/fi/l7j945wij6iru70755syp/Uber-Kenya-gdoc.gdoc?dl=0&rlkey=n5rsg6iqqf8aj7wpvu0de2xev
- **Segment_info.csv** contains info on specific road segments. This includes information on physical characteristics such as the existence of crosswalks or obstacles in the road as well as behavioral characteristics such as people walking along the side of the road, all of which may be associated with the likelihood of a road traffic crash. The columns have been obfuscated but the data may still be useful. It can be linked to physical locations by joining with the geometry in **segments_geometry.geojson**. Some segments have separate rows for each side of the road, and so two rows in **Segment_info.csv** may map to the same road segment in **segments_geometry.geojson**.
- You are also allowed to use data from **movement.uber.com**, and are encouraged to do so. For example, you can get hourly average speeds for different routes from https://movement.uber.com/explore/nairobi/speeds. This data can be mapped to OpenStreetMap ways.

You may NOT use any additional data sources except the ones listed above.

Variable definitions

Train.csv:

- Uid - a unique ID
- Datetime - the date and time a crash occurred
- Latitude and Longitude - the location of the crash (not always exact due to the nature of the data collection)

Sample Submission

- Date - a datetime column (3 hour intervals covering the test period)
- A[N]_Latitude and A[N]_Longitude - used to place ambulance N at a specific location

Weather Info:

- The weather data comes from the GFS dataset. Descriptions of the image bands used to generate this data can be found here.

Segment_info:

- The column headings are obfuscated, but we've maintained relationships: 79_76 and 79_65 are two related questions.

Files available for download

- Train.csv - contains crashes between 2018-01-01 and 2019-06-01, each of which has a location (latitude and longitude) and a time.
- Segment_info.csv: data from road segment surveys
- segments_geometry.geojson: geographical representations of the road segments above
- SampleSubmission.csv: Example of submission format. See the starter notebook for examples creating your own.
- StarterNotebook.ipynb: A basic notebook to show the basics of loading the data, re-creating the scoring method and making your first submission.

Rules

Teams and collaboration

You may participate in competitions as an individual or in a team of up to four people. When [illegible] leaders at team formation. Prizes are transferred only to the individual players or to the team [illegible]

Multiple accounts per user are not permitted, and neither is collaboration or membership across multiple teams. Individuals and their submissions originating from multiple accounts will be immediately disqualified from the platform.

Code must not be shared privately outside of a team. Any code that is shared, must be made available to all competition participants through the platform. (i.e. on the discussion boards).

The Zindi user who sets up a team is the default Team Leader. The Team Leader can invite [illegible] competition rules [illegible] of the competition or the last hour of a hackathon unless otherwise stated in the [illegible]

A team can be disbanded if it has not yet made a submission. Once a submission is made individual members cannot leave the team.

All members in the team receive points associated with their ranking in the competition and there is no split or division of the points between team members.

Datasets and packages

The solution must use publicly-available, open-source packages only.

You may use only the datasets provided for this competition. Automated machine learning tools such as automl are not permitted.

You may use pretrained models as long as they are openly available to everyone.

The data used in this competition is the sole property of Zindi and the competition host. You [illegible] private and 2) doing so does not contravene Zindi's rules of use. [illegible]

You must notify Zindi immediately upon learning of any unauthorised transmission of or unauthorised access to the competition data, and work with Zindi to rectify any unauthorised transmission or access.

Your solution must not infringe the rights of any third party and you must be legally entitled to assign ownership of all rights of copyright in and to the winning solution code to Zindi.

Submissions and winning

You may make a maximum of 10 submissions per day. Your highest-scoring solution on the private leaderboard at the end of the competition will be the one by which you are judged.

Zindi maintains a **public leaderboard** and a **private leaderboard** for each competition. The [illegible] the test dataset, will be made public and will constitute the final ranking for the competition. [illegible] % of

If you are in the top 20 at the time the leaderboard closes, we will email you to request your [illegible] **following the submission guidelines detailed below. Failure to respond will result in disqualification.** [illegible]

If your solution places **1st, 2nd, or 3rd on the final leaderboard**, you will be required to submit [illegible] rights of copyright in and to such winning solution to Zindi. [illegible] agree to assign all worldwide

If two solutions earn identical scores on the leaderboard, the tiebreaker will be the date and time in which the submission was made (the earlier solution will win).

If the error metric requires probabilities to be submitted, do not set thresholds (or round your [illegible] thresholds to their own needs. [illegible] will need the raw probabilities. This will allow the clients to set

The winners will be paid via bank transfer, PayPal, or other international money transfer [illegible] from your bank confirming your banking details. [illegible] proof of identification, proof of residence and a letter

You acknowledge and agree that Zindi may, without any obligation to do so, remove or [illegible] account [illegible] official competition rules. Entry into this competition constitutes your acceptance of these

Zindi is committed to providing solutions of value to our clients and partners. To this end, we [illegible] inherent value of your solution. [illegible] data leaks or any other practices that we deem to compromise the

Zindi also reserves the right to disqualify you and/or your submissions from any competition if [illegible] completely at the discretion of Zindi. [illegible] are irrespective of your position on the leaderboard and

Please refer to the FAQs and Terms of Use for additional rules that may apply to this competition. We reserve the right to update these rules at any time.

- **Reproducibility of submitted code**

    - If your submitted code does not reproduce your score on the leaderboard, we reserve the right to adjust your rank to the score generated by the code you submitted.
    - If your code does not run you will be dropped from the top 10. Please make sure your code runs before submitting your solution.
    - Always set the seed. Rerunning your model should always place you at the same position on the leaderboard. When running your solution, if randomness shifts you down the leaderboard we reserve the right to adjust your rank to the closest score that your submission reproduces.
    - We expect full documentation. This includes:
        - All data used
        - Output data and where they are stored
        - Explanation of features used
        - A requirements file with all packages and versions used
        - Your solution must include the original data provided by Zindi and validated external data (if allowed)
        - All editing of data must be done in a notebook (i.e. not manually in Excel)
        - Environment code to be run. (e.g. Google Colab or the specifications of your local machine)
        - Expected run time for each notebook. This will be useful to the review team for time and resource allocation.

Data standards:

- Your submitted code must run on the original train, test, and other datasets provided.
- If external data is allowed, external data must be freely and publicly available, including pre-trained models with standard libraries. If external data is allowed, any data used should be shared with Zindi to be approved and then shared on the discussion forum. Zindi will also make note of the external data available on the data page.
- Packages:
    - You must submit a requirements file with all packages and versions used.
    - If a requirements file is not provided, solutions will be run on the most recent packages available.
    - Custom packages in your submission notebook will not be accepted.
    - You may only use tools available to everyone i.e. no paid services or free trials that require a credit card.

Consequences of breaking any rules of the competition or submission guidelines:

- **First offence: No prizes or points for 6 months (probation period).** If you are caught cheating, all individuals involved in cheating will be disqualified from the challenge(s) you were caught in and you will be disqualified from winning any competitions or Zindi points for the next six months.

- **Second offence: Banned from the platform.** If you are caught for a second time your Zindi account will be disabled and you will be disqualified from winning any competitions or Zindi points using any other account.
- Teams with individuals who are caught cheating will not be eligible to win prizes or points in the competition in which the cheating occurred, regardless of the individuals' knowledge of or participation in the offence.
- Teams with individuals who have previously committed an offence will not be eligible for any prizes for any competitions during the 6-month probation period.

Monitoring of submissions

- We will review the top 20 solutions of every competition when the competition ends.
- We reserve the right to request code from any user at any time during a challenge. You will have 24 hours to submit your code following the rules for code review (see above). Zindi reserves the right not to explain our reasons for requesting code.
  If you do not submit your code within 24 hours you will be disqualified from winning any competitions or Zindi points for the next six months. If you fall under suspicion again and your code is requested and you fail to submit your code within 24 hours, your Zindi account will be disabled and you will be disqualified from winning any competitions or Zindi points with any other account.