1. MP Races project

Roadraceresults.com

Brief

1. **Search for Canadian MPs who have run in organized races**
2. **Create a ranked list of MPs to pitch ideas to, based on probable responsiveness**

Method

1. Auto-extract MP details (age, riding) with py script which does a google lookup (API based)
2. Cross reference list of sitting MPs with roadraceresults.com database (9M+ results) based on name, age, riding etc.
3. Rank MPs based on number of races run, median finishing times (to determine which individuals are most serious about running)

Interesting problems faced and solutions

P1: Age was not a criterion in the database and had to be extracted from a 'category' string field which took multiple forms (eg. M35-39, F21, F 3035 etc.)

S1: Create new field extracting age (as int) from the field using regex extraction.

P2: Data for finishing times, race distances not clean

S2: Created py scripts to import data directly from MySQL server and add missing data for distance. Time was saved as a string and had unpredictable offsets (such as mins being stored as hours in some instances). Was able to remedy by checking the best time for that race and determining what sort of offset there is (best times are usually lower than a certain time for a particular distance)

2. Race Analytics

Roadraceresults.com

Brief

1. **Create "Race Profiles" for popular races to understand general stats around the race**
2. **Make shareable snippets/ creatives for social media**

Method

1. Run script to group together races which are part of the same series of races (eg: Toronto Goodlife Marathon)
2. Calculate stats such as total kms run, average participation, best times, gender ratio, age distribution etc. for this truncated group of races
3. Export data to excel file which automatically makes a shareable creative
4. The script also automatically exported a list of roadraceresults.com registered users who have run this race in the past

Interesting problems faced and solutions

P1: Race names have changed in during the course of the race's progression over the years, so a direct name lookup yields only a partial list of historical races

S1: Had to creatively use multiple data points including the name, location and "last racecode" field in the database, since none of the these individually yielded complete results

---

P2: Data for cities was user added resulting in multiple typos for city and resulting in fewer points for the map

S2: Used regex to extract nearest match city from the user input value, providing more data points

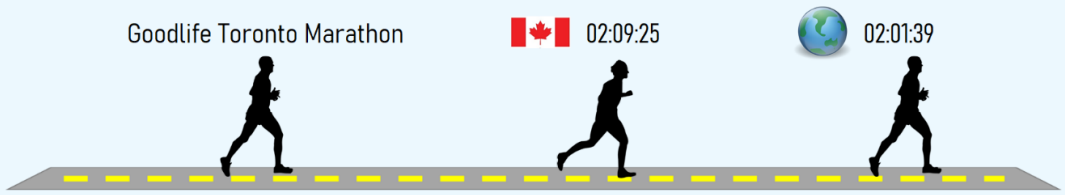
2. Race Analytics Creative

Goodlife Toronto Marathon

1583 on average — Runners participated in the race every year

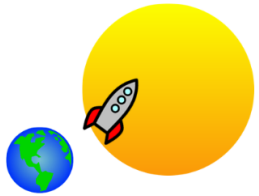
34% females

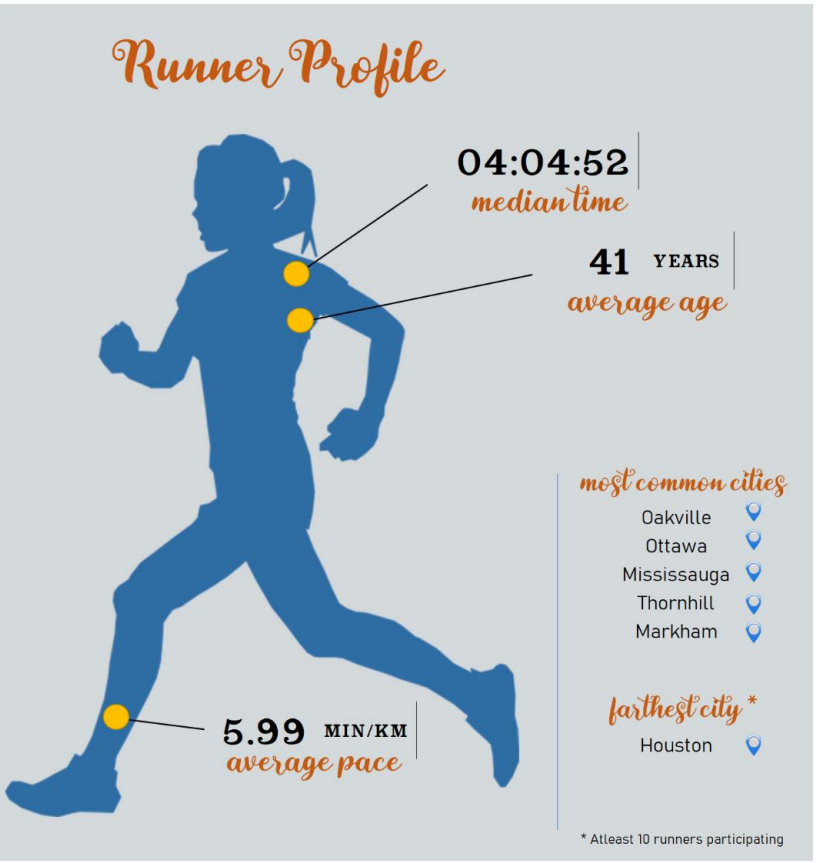
best time...
02:16:42

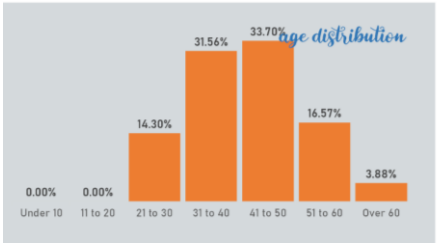


a total of...
1,069,052,520
...kms run...

...enough to make...
4 trips to the Sun and back!







Note: Sensitive, personal and company confidential information is redacted wherever relevant

3. Race Categorization

Roadraceresults.com

Brief

1. **Create a tool to quickly output races which satisfy certain criteria**
2. **Make results exportable to be used for presentations to possible sponsors, race directors etc.**

Method

1. Create front end (Windows) which can lookup the races using sliders for the criteria
2. Use statistical methods to determine the following parameters of the race – seriousness, popularity, professionalism, age distribution, gender distribution
3. Display race list in tool and make export option available.
4. Python libs: tkinter, pandas, numpy, sqlalchemy

Interesting problems faced and solutions

P1: Cleaning data on the fly for large number of races is resource intensive and slow

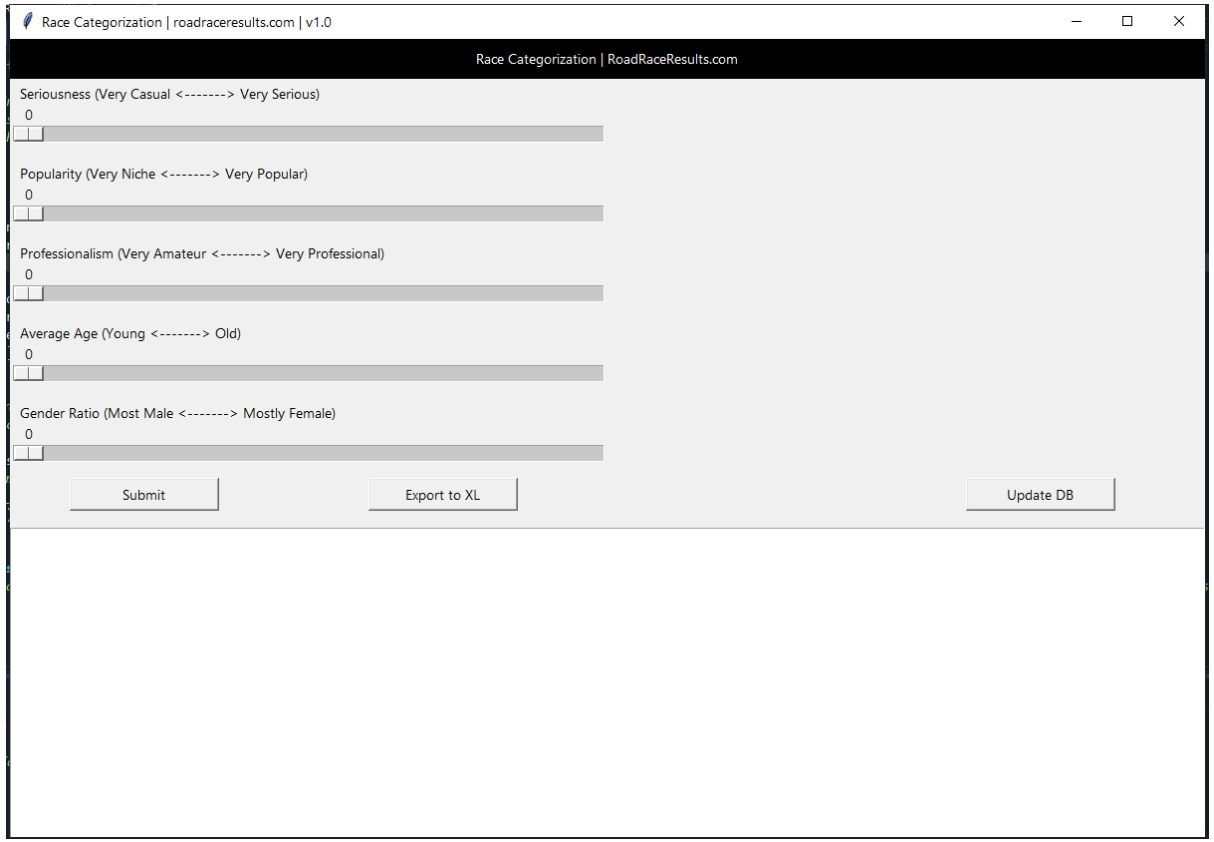
S1: Created local db with "Update" option to speed up outputs

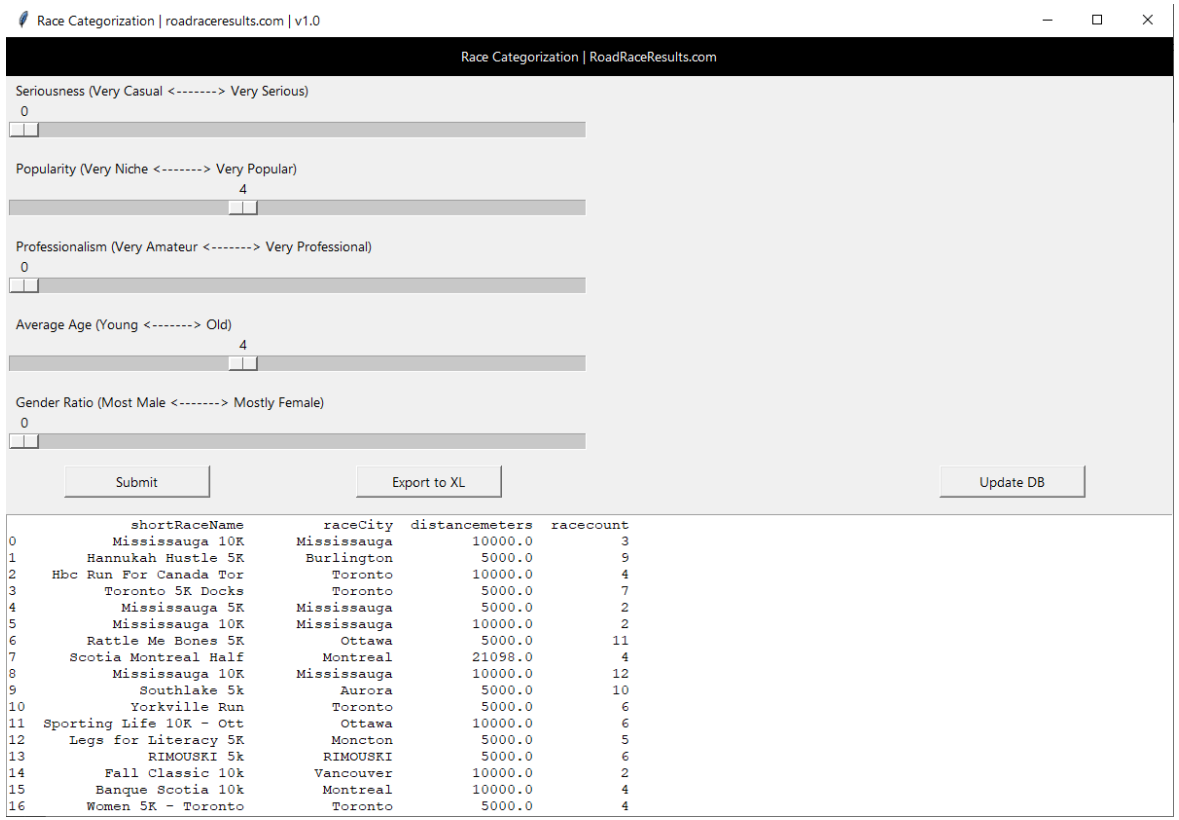
---

P2: Parameters such as seriousness and professionalism are not easily defined.

S2: Created and documented elaborate statistical formula for calculating seriousness (How serious are the runners of this race?) and professionalism (How well do the top performers in this race compare to the accepted benchmarks?)

3. Race Categorization Tool





	shortRaceName	raceCity	distancemeters	racecount
0	Mississauga 10K	Mississauga	10000.0	3
1	Hannukah Hustle 5K	Burlington	5000.0	9
2	Hbc Run For Canada Tor	Toronto	10000.0	4
3	Toronto 5K Docks	Toronto	5000.0	7
4	Mississauga 5K	Mississauga	5000.0	2
5	Mississauga 10K	Mississauga	10000.0	2
6	Rattle Me Bones 5K	Ottawa	5000.0	11
7	Scotia Montreal Half	Montreal	21098.0	4
8	Mississauga 10K	Mississauga	10000.0	12
9	Southlake 5k	Aurora	5000.0	10
10	Yorkville Run	Toronto	5000.0	6
11	Sporting Life 10K - Ott	Ottawa	10000.0	6
12	Legs for Literacy 5K	Moncton	5000.0	5
13	RIMOUSKI 5k	RIMOUSKI	5000.0	6
14	Fall Classic 10k	Vancouver	10000.0	2
15	Banque Scotia 10k	Montreal	10000.0	4
16	Women 5K - Toronto	Toronto	5000.0	4