Competition: DAT102x: Predicting Chronic Hunger

Data Science Capstone

About the Data

Your goal is to predict the annual prevalence of undernourishment at the country level from other socioeconomic indicators. The prevalence of undernourishment expresses "the probability that a randomly selected individual from the population consumes an amount of calories that is insufficient to cover her/his energy requirement for an active and healthy life" (FAOSTAT). It can be understood as the percent of the total population that is facing chronic hunger. Data is compiled from the Food and Agricultural Organization of the United Nations as well as the World Bank.

Target Variable

We're trying to predict the variable `prevalence_of_undernourishment` for each row of the test data set. `prevalence_of_undernourishment` is a positive floating point number (e.g. 1.0) between 0.0 and 100.0 inclusive.

Your job is to:

1. Train a model using the inputs in `train_values.csv` and the labels `train_labels.csv`
2. Predict floats for each row in `test_values.csv` for which you don't know the true prevalence of undernourishment.
3. Output your predictions in a format that matches `submission_format.csv` **exactly**.
4. Upload your predictions to this competition in order to get a score.
5. Export your token and paste it into the assignment grader on edX to get your course grade.

Submission Format

The format for the submission file is two columns with the `row_id` and the `prevalence_of_undernourishment`. The data type of `prevalence_of_undernourishment` is a float, **so make sure there is a decimal point in your submission**. For example `100.0` is a valid float but `100` is *not*.

The first few lines of the `.csv` file that you submit would look like:

```
row_id,prevalence_of_undernourishment
0,1.0
1,1.0
```

```
2,1.0
3,1.0
4,1.0
⋮
```

Performance Metric

We're predicting a numeric quantity, so this is a regression problem. To measure regression, we'll use a metric called root-mean-squared error. It is an error metric, so lower value is better (as opposed to an accuracy metric, where a higher value is better).

RMSE=∑Nn=1(y^n–yn)2)N−−−−−−−−−−−−−√

$$RMSE = \sqrt{\frac{\sum_{n\,=\,1}^{N} (\hat{y}_n - y_n)^2)}{N}}$$

Where $y\hat{\hat{y}}$ is the predicted prevalence of undernourishment and "y" is the actual prevalence of undernourishment. The best possible score is 0, but the worst possible score can be infinite.

Features

There are 45 variables in this dataset. Each row in the dataset represents a country in a given year. Each country has a unique identifier, `country_code`, as country names are not given.

The country codes in the test set **are distinct** from those in the train set. In other words, no country that appears in the train set appears in the test set. Thus, country-specific features (i.e. country dummy variables) will not be an option. However, the countries in the test set still share similar patterns as those in the train set and so other feature engineering will work the same as usual.

When you are doing your local train/test split, it is imperative that you split the data by country so that all years of data for a country appear either in the train set or the test set, but are not split across both. Otherwise, your model will simply interpolate between the known values in your train set rather than learning the underlying patterns in the data. The model will then score poorly when submitted as it will be evaluated on data from entirely different countries.

If your leaderboard score is significantly lower than your local score, it is likely because you have not split your data by country when training your model.

For each country, there are between 5 and 16 years of data in the train set. In the test set, there are at least 10 years of data for each country.

The variables are as follows:

ID

- `country_code` - Unique identifier for each country.
- `year`

Agriculture

- `agricultural_land_area` - Land area in square kilometers that is suitable or used for growing crops or as pastures.
- `percentage_of_arable_land_equipped_for_irrigation` - Percent of total arable land that is equipped for irrigation.
- `cereal_yield` - Average yield in kg/hectare of wheat, rice, maize, barley, oats, rye, millet, sorghum, buckwheat, and mixed grains.
- `droughts_floods_extreme_temps` - Annual average percent of the population that is affected by droughts, floods, or extreme temperature events (average 1990-2009).
- `forest_area` - Land area in square kilometers that is under natural or planted stands of trees in their original location. Excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens.
- `total_land_area` - Total land area of a country in square kilometers.

Demographics

- `fertility_rate` - Number of children that would be born to a woman if she were to live to the end of her childbearing years and bear children in accordance with age-specific fertility rates of the specified year. Measured in births per woman.
- `life_expectancy` - Number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.
- `rural_population` - Number of people living in rural areas.
- `total_population` - Number of people living in a country.
- `urban_population` - Number of people living in urban areas.
- `population_growth` - Annual population growth rate (%).

Economics

- `avg_value_of_food_production` - Estimated food net production value of a country expressed in per capita terms. Measured in constant 2004-06 international dollars per person.
- `cereal_import_dependency_ratio` - The cereal imports dependency ratio tells how much of the available domestic food supply of cereals has been imported and how much comes from the country's own production. It is computed as (cereal imports - cereal exports)/(cereal production + cereal imports - cereal exports) * 100. Negative values indicate that the country is a net exporter of cereals.
- `food_imports_as_share_of_merch_exports` - Value of food imports expressed as a percent of total merchandise exports.
- `gross_domestic_product_per_capita_ppp` - Gross domestic product (GDP) is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is divided by the total population to be expressed in per capita terms. GDP per capita is often used as proxy for average income levels in a country and having it at purchasing power parity (PPP) allows is to be comparable across countries. Measured in constant 2011 international dollars per person.
- `imports_of_goods_and_services` - Value of all goods and other market services received from the rest of the world expressed as a percent of GDP.
- `inequality_index` - The Gini index measures how equal the income distribution is in a country. A Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality.
- `net_oda_received_percent_gni` - Net official development assistance received expressed as a share of gross national income (GNI). The ratio of aid to GNI provides a measure of recipient country's dependency on aid, where higher values indicate a greater dependency.
- `net_oda_received_per_capita` - Net official development assistance received divided by the

total population. Measured in current US$ per capita.

- `tax_revenue_share_gdp` - Tax revenue (value of all taxes collected) expressed as a percent of GDP.
- `trade_in_services` - Trade in services (sum of service exports and imports) expressed as a percent of GDP.
- `per_capita_food_production_variability` - Variability of food production value (avg_value_of_food_production). Measured in constant 2004-06 international dollars per person.
- `per_capita_food_supply_variability` - Variability of the food supply in per capita terms. Measured in kcal/capita/day per person.

Education

- `adult_literacy_rate` - Percent of people ages 15 and above who can both read, write and understand a short simple statement about their everyday life.
- `school_enrollment_rate_female` - Percent of female primary education-aged children enrolled in school.
- `school_enrollment_rate_total` - Percent of all primary education-aged children enrolled in school.

Food security

- `avg_supply_of_protein_of_animal_origin` - Average protein supply expressed in grams per capita per day. It includes protein from meat, milk, eggs, fish, seafood, and other animal products.
- `caloric_energy_from_cereals_roots_tubers` - Percent of total dietary energy supply coming from cereals, roots and tubers.

Health

- `access_to_improved_sanitation` - Percent of the population with at least adequate access to excreta disposal facilities that can effectively prevent human, animal, and insect contact with excreta. Improved facilities range from simple but protected pit latrines to flush toilets with a sewerage connection.
- `access_to_improved_water_sources` - Percent of the population with reasonable access to an adequate amount of water from an improved source, such as a household connection, public standpipe, borehole, protected well or spring, and rainwater collection.
- `anemia_prevalence` - Percent of women of reproductive age (15-49 years) who meet the clinical definition of anemia.
- `obesity_prevalence` - Percent of adults ages 18 and over whose Body Mass Index is more than 30 kg/m2.
- `open_defecation` - Percent of the population defecating in the open, such as in fields, forest, bushes, open bodies of water, or on beaches.
- `hiv_incidence` - Number of new HIV infections among uninfected populations ages 15-49 expressed per 100 people in the uninfected population in the previous year.

Infrastructure

- `rail_lines_density` - Ratio between the length of railway route available for train service and the area of the country (per 100 sq km of land area).
- `access_to_electricity` - Percent of population with access to electricity.

- `co2_emissions` - Carbon dioxide emissions in kt (thousands of metric tons).

Labor

- `unemployment_rate` - Percent of the labor force that is without work but available for and seeking employment.
- `total_labor_force` - Total number of people who are currently employed, people who are unemployed but seeking work, and first-time job-seekers.

Politics

- `military_expenditure_share_gdp` - Spending on the armed forces and defense ministries expressed as a percent of the country's GDP.
- `proportion_of_seats_held_by_women_in_gov` - Percent of seats in national parliaments held by women.
- `political_stability` - Index of the perceived likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including politically-motivated violence and terrorism.

Example Row

Here's an example of one of the rows in the dataset so that you can see the kinds of values you might expect in the dataset. Most are numeric, one is categorical, and there can be missing values.

	0
country_code	889f053
year	2002
agricultural_land_area	235078
percentage_of_arable_land_equipped_for_irrigation	38.5585
cereal_yield	935.754
droughts_floods_extreme_temps	NaN
forest_area	5397.74
total_land_area	537408
fertility_rate	5.92898
life_expectancy	60.4522
rural_population	1.35409e+07
total_population	1.88995e+07
urban_population	5.08645e+06
population_growth	2.86493
avg_value_of_food_production	60.8913
cereal_import_dependency_ratio	78.4528
food_imports_as_share_of_merch_exports	22.0805
gross_domestic_product_per_capita_ppp	3969.52
imports_of_goods_and_services	38.3518
inequality_index	NaN
net_oda_received_percent_gni	2.23356
net_oda_received_per_capita	11.659

	o
tax_revenue_share_gdp	NaN
trade_in_services	NaN
per_capita_food_production_variability	1.96809
per_capita_food_supply_variability	15.6935
adult_literacy_rate	NaN
school_enrollment_rate_female	NaN
school_enrollment_rate_total	NaN
avg_supply_of_protein_of_animal_origin	11.0241
caloric_energy_from_cereals_roots_tubers	64.1296
access_to_improved_sanitation	43.2865
access_to_improved_water_sources	57.9328
anemia_prevalence	59.0614
obesity_prevalence	8.29394
open_defecation	28.9335
hiv_incidence	0.00997508
rail_lines_density	NaN
access_to_electricity	52.4332
co2_emissions	15485.1
unemployment_rate	14.788
total_labor_force	4.35052e+06
military_expenditure_share_gdp	7.02107
proportion_of_seats_held_by_women_in_gov	0.698153
political_stability	-1.3938

References

- FAOSTAT, Food and Agriculture Organization of the United Nations. Definitions and Standards. http://www.fao.org/faostat/en/.

One account per participant

You cannot sign up to this website http://datasciencecapstone.org ("Website") from multiple accounts and therefore you cannot submit from multiple accounts.

Private sharing of code

Privately sharing code or data outside of teams is not permitted.

Public dissemination of entries

DrivenData and the competition host have the right to publicly disseminate any entries or models.

Documentation for winning solutions

Winning solutions must be documented using the provided Winning Model Documentation Template in order to be eligible for recognition and prize money if offered.

Submission limits

The number of submissions per day is restricted to a fixed value on a per-competition basis. Any attempt to circumvent this limit will result in disqualification.

External data

External data is not allowed unless otherwise noted explicitly on these competition pages. Participants agree to make no attempt to use additional data or data sources not provided on these competition pages.

These are the complete, official rules for the Competition (the **'Competition Rules'**) and incorporate by reference the contents of the Competition Website listed above.

By downloading a dataset linked from the Competition Website, submitting an entry to this Competition, or joining a Team in this Competition, you are agreeing to be bound by these Competition Rules which constitute a binding agreement between you and DrivenData and, if applicable, any rules and restrictions that may be imposed by the Competition Sponsor.

The Competition is sponsored by the Competition Sponsor listed above and hosted on the Competition Sponsor's behalf by DrivenData, Inc (**'DrivenData'**). The Competition will run according to the dates listed on the Competition Website. From time to time, DrivenData may sponsor a Competition, in which case DrivenData will act as the Competition Sponsor.

Each registered individual or Team is referred to as a Participant. You may only compete using a single, unique Website account registered at the Website. Competing using more than one account per individual is a breach of these Competition Rules and Competition Sponsor reserves the right to disqualify any individual (or Team including an individual) who is found to breach these Competition Rules.

Eligibility

The Competition is open to all individuals over the age of 18 at the time of entry and to all validly formed legal entities that have not declared or been declared in bankruptcy.

Officers, directors, employees and advisory board members (and their immediate families and members of the same household) of the Competition Sponsor, DrivenData and their respective affiliates, subsidiaries, contractors, agents, judges and advertising and promotion agencies are not eligible to participate in the Competition.

You are not eligible to receive any Prize in the Competition if you are a resident of a country designated by the United States Treasury's Office of Foreign Assets Control.

Registration

If you meet the eligibility requirements and would like to participate, then you must first complete the registration process at the Website within the Competition Period. After you complete the registration process, you will receive access to the Data (described on the Competition page) that will enable you to develop and submit one or more Entries (as defined below). All Entries must be received during the Competition Period. To register, visit the Website and follow the onscreen instructions to complete and submit your registration. All of the registration information that you provide is collectively referred to as your "Account". (If you have already created an Account at the Website, enter your user name and password and follow the on-screen instructions).

After you register individually, you may join a team (a "Team"), only if expressly permitted by these rules below, but you may register only once. If you register for the Competition more than once, you will be, and the remainder of your Team may be, disqualified in Sponsor's sole discretion.

You acknowledge and agree that you are solely responsible for abiding by your employer's policies regarding participation in the Competition. Sponsor disclaims any and all liability or responsibility for disputes arising between an employee and employer related to this Competition.

Once you have completed the registration process, you or your Team will be provided with access to the Data that you use to develop your Entries.

BY ENTERING THE COMPETITION YOU ACCEPT THE CONDITIONS STATED IN THESE OFFICIAL RULES INCLUDING THE RULES ON THE COMPETITION PAGE, AGREE TO BE BOUND BY THE DECISIONS OF THE JUDGES AND WARRANT THAT YOU ARE ELIGIBLE TO PARTICIPATE IN THE COMPETITION. IF YOU DO NOT ACCEPT ALL OF THESE OFFICIAL RULES, THEN PLEASE DO NOT REGISTER FOR THE COMPETITION. WE RECOMMEND THAT YOU PRINT OUT A COPY OF THESE OFFICIAL RULES FOR YOUR FUTURE REFERENCE.

Teams

FORMING A TEAM. If DrivenData enables team functionality for this Competition, multiple individuals or entities may collaborate as a team (**"Team"**). You may not participate on more than one Team. Each Team member must be a single individual operating a separate Website account. You must register individually for the Competition before joining a Team. You must confirm your Team membership to make it official by responding to the Team notification message which will be sent to your Account.

TEAM PRIZES. If a Team wins a monetary Prize, Competition Sponsor will allocate the Prize money in even shares between the Team members unless the Team unanimously contacts DrivenData within three business days following the Submission deadline to request an alternative prize distribution.

Entry submissions

Each Entry must be uploaded to the Website in the manner and format specified on the Website. All Entries must be received during the Competition Period. An **"Entry"** is the data submitted in the manner and format specified on the Website via the Entry form. The number of Entries a Participant may submit during each calendar day of the Competition Period will be displayed on the Competition page. Sponsor reserves the right to request that a Participant submit the algorithm associated with an Entry to Sponsor. If the Competition is a multi-stage Competition with temporally separate training data and/or leaderboard data, one or more valid Entries must be submitted and selected during each stage of the Competition in the manner described on the

Website.

USER SUBMISSIONS MUST IN THEIR ENTIRETY COMPLY WITH ALL APPLICABLE FEDERAL, STATE, LOCAL, AND INTERNATIONAL LAWS AND REGULATIONS. WITHOUT LIMITING THE FOREGOING, IN ORDER TO BE ELIGIBLE, YOU WARRANT THAT (UNLESS OTHERWISE SPECIFIED IN THE PROBLEM STATEMENT) YOUR SUBMISSION FOR THIS COMPETITION:

- Does not include or anticipate the inclusion of any content that is in violation of or infringes third party intellectual property rights including, but not limited to copyrights, including music copyrights, patents, trade secrets, trademarks, and rights of publicity or privacy.
- Is free and clear of all liens, claims, encumbrances or demands of any third parties.
- Does not include or anticipate the inclusion of any unsuitable or offensive content, including nudity, sexually explicit, disparaging, libelous or other inappropriate content.
- Has not been entered in previous challenges, or won previous awards.
- Has not been published or distributed previously in any media.
- Is suitable for a general audience.
- Does not contain any claims that are not and cannot be substantiated or that would be false and/or misleading to a reasonable consumer.

Selection of winners

This Competition is a challenge of skill and the final results are determined solely by leaderboard ranking on the private leaderboard (subject to compliance with these Competition Rules). Participants' scores and ranks on the Competition Website at any given stage of the Competition will be based on the evaluation metric described on the Competition Website, as determined by applying the predictions in the Submission to the ground truth of a validation dataset whose instances were a fixed set sampled from the Data.

The evaluation metric used for scoring and ranking Submissions will be displayed on the Competition Website.

Any prize awards are subject to verification of eligibility and compliance with these Competition Rules. All decisions of DrivenData, the Competition Sponsor and/or judges will be final and binding on all matters relating to this Competition. DrivenData and Competition Sponsor reserves the right to examine the Submission and any associated code or documentation for compliance with these Competition Rules. In the event that the Submission demonstrates a breach of these Competition Rules, Competition Sponsor may at its discretion take either of the following actions:

- disqualify your Submission(s); or
- require that you remediate within one week all issues identified in your Submission(s) (including, without limitation, the resolution of license conflicts, the fulfillment of all obligations required by software licenses, and the removal of any software that violates the software restrictions).

A tie between two or more valid and identically ranked submissions will be resolved in favor of the tied submission that was submitted first.

A Participant may decline to be nominated as a Winner by notifying DrivenData directly within one week following the Competition deadline, in which case the declining Participant forgoes any prize or other features associated with winning the Competition. DrivenData reserves the right to disqualify a Participant who so declines at DrivenData's sole discretion if DrivenData deems disqualification appropriate.

Prizes and conditions

NO PRIZES

There are no prizes being offered in association with challenges hosted on datasciencecapstone.org.

CHEATING. Participating using more than one Website account is deemed cheating and, if discovered, will result in disqualification from the Competition and any other affected Competitions and may result in banning or deactivation of affected Website accounts.

DrivenData reserves the right to request information associated with our investigation of suspected cheating. Failure to respond to these requests (including failure to furnish the requested information) within ten (10) days is grounds for disqualification.

Data use and code sharing

DATA. "Data" means the Data or Datasets linked from the Competition Website for the purpose of use by Participants in the Competition. For the avoidance of doubt, Data is deemed for the purpose of these Competition Rules to include any prototype or executable code provided to Participants by DrivenData or Competition Sponsor via the Website. Participants must use the Data only as permitted by these Competition Rules and any associated data use rules specified on the Competition Website.

Unless otherwise permitted by the terms of the Competition Website, Participants must use the Data solely for the purpose and duration of the Competition, including but not limited to reading and learning from the Data, analyzing the Data, modifying the Data and generally preparing your Submission and any underlying models and participating in forum discussions on the Website. Participants agree to use suitable measures to prevent persons who have not formally agreed to these Competition Rules from gaining access to the Data and agree not to transmit, duplicate, publish, redistribute or otherwise provide or make available the Data to any party not participating in the Competition. Participants agree to notify DrivenData immediately upon learning of any possible unauthorized transmission or unauthorized access of the Data and agree to work with DrivenData to rectify any unauthorized transmission. Participants agree that participation in the Competition shall not be construed as having or being granted a license (expressly, by implication, estoppel, or otherwise) under, or any right of ownership in, any of the Data.

EXTERNAL DATA. Unless otherwise expressly stated on the Competition Website, Participants must not use data other than the Data to develop and test their models and Submissions. Competition Sponsor reserves the right in its sole discretion to disqualify any Participant who Competition Sponsor discovers has undertaken or attempted to undertake the use of data other than the Data, or who uses the Data other than as permitted according to the Competition Website and in these Competition Rules, in the course of the Competition.

CODE SHARING. Participants are prohibited from privately sharing source or executable code developed in connection with or based upon the Data, and any such sharing is a breach of these Competition Rules and may result in disqualification.

Participant warranties and obligations

By registering, you agree that (a) your Account is complete, correct and accurate and (b) your registration may be rejected or terminated and all Entries submitted by you and/or your Team may

be disqualified if any of the information in your Account is (or Competition Sponsor has reasonable grounds to believe it is) incomplete, incorrect or inaccurate. You are solely responsible for your Account. All registration information is deemed collected in the United States.

Participation is subject to all federal, state and local laws and regulations. Void where prohibited or restricted by law. You are responsible for checking applicable laws and regulations in your jurisdiction before participating in the Competition to make sure that your participation is legal. You are responsible for all taxes and reporting related to any award that you may receive as part of the Competition. You are responsible for abiding by your employer's policies regarding participation in the Competition. Competition Sponsor disclaims any and all liability or responsibility for disputes arising between you and your employer related to this Competition.

Each Participant is solely responsible for all equipment, including but not necessarily limited to a computer and internet connection necessary to access the Website and to develop and upload any Submission, and any telephone, data, hosting or other service fees associated with such access, as well as all costs incurred by or behalf of the Entrant in participating in the Competition.

By entering a Submission, you represent and warrant that all information you enter on the Website is true and complete to the best of your knowledge, that you have the right and authority to make the Submission (including any underlying code and model) on your own behalf or on behalf of the persons and entities that you specify within the Submission, and that your Submission:

- complies with all applicable federal, state, local and international laws and regulations;
- is your own original work, or is used by permission, in which case full and proper credit and identify is given and the third party contributions are clearly identified within your Submission;
- does not contain confidential information or trade secrets and is not the subject of a registered patent or pending patent application;
- does not violate or infringe upon the patent rights, industrial design rights, copyrights, trademarks, rights of privacy, publicity or other intellectual property or other rights of any person or entity;
- does not contain malicious code, such as viruses, timebombs, cancelbots, worms, Trojan horses or other potentially harmful programs or other material or information;
- does not and will not violate any applicable law, statute, ordinance, rule or regulation;
- does not trigger any reporting or royalty obligation to any third party; and
- was not previously published and has not won any other prize/award.

A breach of any of these warranties will result in the corresponding Submission being invalid.

Confidentiality

Confidential Information Defined. As used in this Section, "Confidential Information" means the Competition Sponsor Materials, and any information provided by DrivenData hereunder, whether of a technical, business, or other nature (including, without limitation, information relating to a party's technology, software, products, services, designs, methodologies, business plans, finances, marketing plans, Sponsors, prospects, or other affairs), that is received by Participant.

Exclusions. Participant's obligations under this Section as to the disclosing party's Confidential Information does not include any information that Participant can document: (a) was known to Participant prior to receiving the same from the disclosing party in connection with this Agreement; (b) is independently developed by Participant without use of or reference to the Confidential Information of the disclosing party; (c) is acquired by Participant from another source

without restriction as to use or disclosure; or (d) is or becomes part of the public domain through no fault or action of Participant.

Nondisclosure. In consideration of the terms and conditions of this Agreement, and for other good and valuable consideration, the receipt and sufficiency of which is acknowledged, during and after the term of this Agreement Participant will: (a) use the disclosing party's Confidential Information solely for the purpose for which it is provided; (b) not disclose the disclosing party's Confidential Information to a third party unless the third party must access the Confidential Information to perform in accordance with this Agreement and the third party has executed a written agreement that contains terms that are substantially similar to the terms contained in this Section; and (c) maintain the secrecy of, and protect from unauthorized use and disclosure, the disclosing party's Confidential Information to the same extent (but using no less than a reasonable degree of care) that Participant protects its own Confidential Information of a similar nature. If Participant is required by law or by any governmental or semi-governmental agency or court to disclose the disclosing party's Confidential Information or the terms of this Agreement Participant must give prompt written notice of such requirement to the disclosing party before such disclosure and assist the disclosing party in attempting to obtain an order protecting the Confidential Information from public disclosure.

Return of Information. Upon request, Participant will destroy or deliver to the disclosing party all Confidential Information Participant may have in its possession or control.

Injunctive Relief. Participant acknowledges that any violation or threatened violation of this Section may cause irreparable injury to the other party, entitling the disclosing party to obtain injunctive relief in addition to all legal remedies.

Equipment and costs of participation

Each Participant is solely responsible for all equipment, including a computer and modem, necessary to establish a connection to the World Wide Web, access to the World Wide Web and any telephone, data, hosting or other service fees associated with such access, as well as all costs incurred by or behalf of the Participant in participating in the Competition.

Delivery and receipt of entries

Sponsor is not responsible for (a) late, lost, stolen, damaged, garbled, incomplete, incorrect or misdirected Entries or other communications, (b) errors, omissions, interruptions, deletions, defects, or delays in operations or transmission of information, in each case whether arising by way of technical or other failures or malfunctions of computer hardware, software, communications devices, or transmission lines, or (c) data corruption, theft, destruction, unauthorized access to or alteration of Entry materials, loss or otherwise. Sponsor is not responsible for electronic communications or emails which are undeliverable as a result of any form of active or passive filtering of any kind, or insufficient space in any email account to receive email messages. Sponsor disclaims any liability for damage to any computer system resulting from participation in, or accessing or downloading information in connection with, the Competition.

Reservation of rights

DrivenData, in its sole discretion, reserves the right to disqualify any person tampering with the entry process, the operation of the Web site, the competition process, or is otherwise in violation of the rules. DrivenData reserves the right at any time to disqualify a Submission from a Competition

where, acting in good faith, it believes there are reasonable grounds to warrant disqualification. For example, DrivenData determination that the Submission does not provide the functionality described or required, or the Submission appears to be purposely designed to circumvent these Rules or the spirit of the Competition would be grounds for disqualification. DrivenData reserves the right to cancel, terminate or modify the competition if it is not capable of completion as planned for any reason, including infection by computer virus, bugs, tampering, unauthorized intervention or technical failures of any sort.

General release

By entering the Competition, you release and discharge the Competition Entities from any liability whatsoever in connection with the Competition or with the acceptance, possession, use or misuse of any prize including, without limitation, legal claims, costs, injuries, losses or damages, demands or actions of any kind (including, without limitation: personal injuries; death; damage to, loss or destruction of property; rights of publicity or privacy; and defamation or portrayal in a false light). The Competition Entities will not be responsible for typographical, printing or other inadvertent errors in these Official Rules or in other materials relating to the Competition. Additionally, you hereby agree to indemnify the Competition Entities from any and all losses, damages, costs, expenses, rights, claims, demands and actions (including attorney's fees and expenses for litigation and settlement), which may be brought against any one or more of them by anyone claiming to have suffered loss or damage as a result of your participation in the Competition.

Limitations of liability

By participating in the Competition, each Participant agrees that: (1) any and all disputes, claims, and causes of action arising out of or in connection with the Competition, or any prize awarded, shall be resolved individually without resort to any form of class action; (2) any claims, judgments and awards shall be limited to actual out-of-pocket costs incurred, including costs associated with entering the Competition, but in no event attorney's fees; and (3) under no circumstances will a Participant be permitted to obtain any award for, and each Participant hereby waives all rights to claim, punitive, incidental or consequential damages and any and all rights to have damages multiplied or otherwise increased and any other damages, other than damages for actual out-of-pocket expenses.

All federal, state, and local laws apply. DrivenData, Inc. and the employees, officers, directors, shareholders, agents, representatives, affiliates, subsidiaries, and advertising, promotional and legal advisors are not responsible and shall not be liable for: (i) late, lost, delayed, damaged, misdirected, incomplete, illegible, or unintelligible entries; (ii) telephone, electronic, hardware, or software program, network, Internet, or computer malfunctions, failures, or difficulties; (iii) errors in transmission; (iv) any condition caused by events beyond the control of DrivenData, Inc. that may cause the contest to be disrupted or corrupted; (v) any injuries, losses, or damages of any kind caused by a prize or resulting from acceptance, possession, or use of a prize, or from participation in the contest; or (vi) any printing or typographical errors in any materials associated with the contest.

Disclaimer of Warranties

You understand that we cannot and do not guarantee or warrant that files available for downloading from the internet or the Website will be free of viruses or other destructive code. You are responsible for implementing sufficient procedures and checkpoints to satisfy your particular requirements for anti-virus protection and accuracy of data input and output, and for maintaining a

means external to our site for any reconstruction of any lost data. WE WILL NOT BE LIABLE FOR ANY LOSS OR DAMAGE CAUSED BY A DISTRIBUTED DENIAL-OF-SERVICE ATTACK, VIRUSES OR OTHER TECHNOLOGICALLY HARMFUL MATERIAL THAT MAY INFECT YOUR COMPUTER EQUIPMENT, COMPUTER PROGRAMS, DATA OR OTHER PROPRIETARY MATERIAL DUE TO YOUR USE OF THE WEBSITE OR ANY SERVICES OR ITEMS OBTAINED THROUGH THE WEBSITE OR TO YOUR DOWNLOADING OF ANY MATERIAL POSTED ON IT, OR ON ANY WEBSITE LINKED TO IT.

YOUR USE OF THE WEBSITE, ITS CONTENT AND ANY SERVICES OR ITEMS OBTAINED THROUGH THE WEBSITE IS AT YOUR OWN RISK. THE WEBSITE, ITS CONTENT AND ANY SERVICES OR ITEMS OBTAINED THROUGH THE WEBSITE ARE PROVIDED ON AN "AS IS" AND "AS AVAILABLE" BASIS, WITHOUT ANY WARRANTIES OF ANY KIND, EITHER EXPRESS OR IMPLIED. NEITHER THE COMPANY NOR ANY PERSON ASSOCIATED WITH THE COMPANY MAKES ANY WARRANTY OR REPRESENTATION WITH RESPECT TO THE COMPLETENESS, SECURITY, RELIABILITY, QUALITY, ACCURACY OR AVAILABILITY OF THE WEBSITE. WITHOUT LIMITING THE FOREGOING, NEITHER THE COMPANY NOR ANYONE ASSOCIATED WITH THE COMPANY REPRESENTS OR WARRANTS THAT THE WEBSITE, ITS CONTENT OR ANY SERVICES OR ITEMS OBTAINED THROUGH THE WEBSITE WILL BE ACCURATE, RELIABLE, ERROR-FREE OR UNINTERRUPTED, THAT DEFECTS WILL BE CORRECTED, THAT OUR SITE OR THE SERVER THAT MAKES IT AVAILABLE ARE FREE OF VIRUSES OR OTHER HARMFUL COMPONENTS OR THAT THE WEBSITE OR ANY SERVICES OR ITEMS OBTAINED THROUGH THE WEBSITE WILL OTHERWISE MEET YOUR NEEDS OR EXPECTATIONS.

THE COMPANY HEREBY DISCLAIMS ALL WARRANTIES OF ANY KIND, WHETHER EXPRESS OR IMPLIED, STATUTORY OR OTHERWISE, INCLUDING BUT NOT LIMITED TO ANY WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT AND FITNESS FOR PARTICULAR PURPOSE. THE FOREGOING DOES NOT AFFECT ANY WARRANTIES WHICH CANNOT BE EXCLUDED OR LIMITED UNDER APPLICABLE LAW.

Governing Law and Jurisdiction

DrivenData shall be the sole interpreter of these Official Rules.

All matters relating to the Official Rules or the Competition and any dispute or claim arising therefrom or related thereto (in each case, including non-contractual disputes or claims), shall be governed by and construed in accordance with the internal laws of the Commonwealth of Massachusetts without giving effect to any choice or conflict of law provision or rule (whether of the Commonwealth of Massachusetts or any other jurisdiction).

Any legal suit, action or proceeding arising out of, or related to, the Competition or the Website shall be instituted exclusively in the federal courts of the United States or the courts of the Commonwealth of Massachusetts although we retain the right to bring any suit, action or proceeding against you for breach of these Terms of Use in your country of residence or any other relevant country. You waive any and all objections to the exercise of jurisdiction over you by such courts and to venue in such courts.

The invalidity or unenforceability of any provision of these Official Rules shall not affect the validity or enforceability of any other provision. In the event that any provision is determined to be invalid or otherwise unenforceable or illegal, these Official Rules shall otherwise remain in effect and be construed in accordance with their terms as if the invalid or illegal provision was not contained herein.