**Smart-Hack Hackathon**

**DISCLAIMER: Make sure you read all the instructions- right until the end of this page!**

**We are doing a hackathon with Shopify data, focused on marketing insights**. You’ve been provided with two data sets from two different vendors. You should use those data sets to develop your models and upload to the platform. The method you develop should be able to function on various datasets, including ones that you haven’t seen before.

**Prizes**

1st Prize: $2,000 CASH

2nd Prize: $1,000 CASH

3rd Prize: $500

4th Prize: $250

**How to get started**

[**Watch this demo**](https://www.smarter.ai/blog/hackathon-experiment).

* If you’ve already registered as a creator-login in, if you don’t have an account- [**get one here**](https://www.smarter.ai/become-a-creator-new).
* You have been provided with two different data sets from two different sources. Access one [**HERE**](https://drive.google.com/file/d/1yM57sXdKrSfiHP5YxVNnhRoz-QUaeC6S/view?usp=sharing) and one [**HERE**](https://drive.google.com/file/d/1Ga0Id_qt4WsX3P06yee-tRN9rYbYXCQX/view?usp=sharing). But please note- there will be a held-out test set for evaluation which you cannot access.
* Here are some demos to help you get started:
  + [How to start the hackathon experiment](https://www.smarter.ai/blog/hackathon-experiment)
  + [How to create an experiment](https://www.smarter.ai/blog/create-your-first-experiment-demo)
  + [How to upload and read data](https://www.smarter.ai/blog/how-to-upload-and-read-data)
  + [How to make your components communicate](https://www.smarter.ai/blog/communication-between-components)

**Data source**: Standard Shopify Data. *Shopify is a subscription-based software that allows anyone to set up an online store and sell their products.*

**Target**: Any insights that will optimize marketing strategy

**Available features**: The data is a mix of Shopify 'Acquisition' and 'Behavior' reports and contains information about where website visitors come from (e.g. country or referrer) as well as checkouts and purchases. It is comprised of the following website visitor features:

* page\_type: Landing page
* location\_country: Country of origin
* referrer\_host: URL of referrer
* referrer\_source: Type of referral (Search, Direct, Email, Social)
* referrer\_name: Name of the referral source
* ua\_os: Device operating system
* ua\_form\_factor: Device Type
* hour: The timestamp of the visit (data is aggregated over 2h windows)
* total\_sessions: The number of sessions during the specified 2h window. A session is a website visit of 30min or less. If someone stays longer on the website, they create a new session.
* total\_checkouts: The number of sessions during the specified 2h window.
* total\_orders\_placed: The number of sessions during the specified 2h window.
* total\_carts: The number of items put in carts during the specified 2h window
* total\_pageviews: The number of pages viewed during all sessions in the 2h window
* total\_conversion: The conversion rate (total\_orders\_placed / total\_sessions) during the specified 2h window.
* total\_bounce\_rate: The number of sessions resulting in no further actions (pageviews, carts, checkouts, purchase) during the specified 2h window.

If referrer information is unknown, this is either due to traffic being direct (URL typed in / bookmarked) or tracking being disabled by the website visitor.

**Rules**

* It can include machine learning but doesn’t *have* to
* Has to be written in Python
* It has to work within[Smarter.ai](http://smarter.ai/)
* Email the version slug/url to email address *hackathon@smarter.ai*
* Maximum submissions of three solutions per person

Judging parameters

**MOST IMPORTANT**: Does it work on a data set you have not seen?

We are evaluating who can find the *deepest* insight from Shopify data sets.

Summary report maximum 500 words, why do you think this is useful, what does it do?

**Performance**

* Model accuracy
* Compute and time requirements
* F1 Score
* Precision
* ROC Curve
* R2 coefficient
* Mean squared error MSE
* Mean absolute error MAE
* Time series
* Mean average percentage error
* Symmetric mean average percentage error

**Usability**

* Interface presentation
* Ease of use
* Business value
* How easily will the customer understand what this will do for them and what they will get out of it

**BONUS**

* If you have machine learning it would be beneficial to have interpretability, why is the model predicting what it is predicting
* If you have machine learning it would be beneficial to have uncertainty estimation: How certain is the model in its predictions.

**Schedule**

|  |  |
| --- | --- |
| **Action** | **Date** |
| Launch | 18 Feb |
| Submissions closed | 10 March |
| Judging commences | 11 March |
| Winner announced | 16 March |

Ready, steady... GO!