



(/contest/mckinsey-
analytics-online-
hackathon-4/)
submissions)



McKinsey Analytics Online Hackathon

July 20 - 22, 2018



Registered

Starts at	Fri Jul 20 2018 05:30:00 GMT+0530 (India Standard Time)
Closes on	Mon Jul 23 2018 05:30:00 GMT+0530 (India Standard Time)
Mode	Online
Fee	Free
Location	Online
# Participants	4694
Prizes	NIPS conference ticket* + 5,000 USD

Contest ends in 1 days: 19 hours: 36 minutes: 50 seconds

About McKinsey Analytics Online Hackathon

During our online Hackathon, you will get a glimpse at the sort of problems and challenges that our McKinsey data scientists solve on a daily basis. The winner will receive a NIPS conference ticket and 5,000 USD for personal expenses. The best participants may also be invited for interviews with McKinsey Analytics.

About McKinsey & Company

McKinsey & Company is a global management consulting firm, deeply committed to helping institutions in the private, public and social sectors achieve lasting success. For over nine decades, our primary objective has been to serve as our clients' most trusted external advisor. With consultants in over 120 locations, in over 60 countries, across industries and functions, we bring unparalleled expertise to clients anywhere in the world. We work closely with teams at all levels of an organization to shape winning strategies, mobilize for change, build capabilities and drive successful execution.

McKinsey Analytics helps clients achieve better performance through data. We work together with clients to build analytics-driven organizations, providing end-to-end support covering strategy, operations, data science, implementation and change management. Our engagements range from use-case specific applications to full-scale analytics transformations. Our teams of consultants, data scientists, and engineers work together with clients to identify opportunities, assess available data, define solutions, establish optimal hosting environments, ingest data, develop cutting-edge algorithms, visualize outputs, and assess impact while building capabilities to sustain and expand it. Learn more at <http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights> (<http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights>)

Job Description - Data Scientist

Who You'll Work With

You'll work as part of the McKinsey Analytics team in one of our offices worldwide, for instance Waltham, MA, Chicago, New York, Sao Paulo, Madrid, London, Brussels, Dusseldorf, Milan, Wroclaw, Moscow, Gurgaon, Chennai, Bangalore, yet other locations are also available. You will work with data engineers and data translators to deliver the most cutting-edge analytics solutions to our Clients from various industries. Through the combination of strategic insights and advanced analytics technologies, you will be solving the most critical problems leading global organizations face.

What You'll Do

Upon joining McKinsey Analytics, you will have a chance to work and learn from the best in class data scientists, data engineers and analytics translators. Here are some examples of projects you could be engaged in:

- Algorithmic route optimization for revenue improvement of a logistics company.
- Applying analytics in financial institutions' fight against fraud.



Feedback

- Simulating mining operations for our American gold mining client that identified opportunities to raise capacity by 20-40 percent.
- Using big data to build better predictive models around individual patients to better diagnose and treat disease.

Qualifications

- Experience in statistical modelling and/or machine learning.
- Experience in applying data science methods to business problems.
- Programming experience in at least 2 of the following language: R, Python, Scala, SQL.
- Educated to MSc or PhD level in the field of Computer Science, Machine Learning, Applied Statistics, Mathematics.
- Strong presentation and communication skills, with a knack for explaining complex analytical concepts to people from other fields.
- Team leadership, mentoring and project management skills.

Prizes:

NIPS conference ticket with flights & accomodation + 5,000 USD

Rules of Participation

McKinsey Analytics Hackathon Rules

1. The contest starts Friday, 20th July 2018, 00:00:00 (GMT) . The data set and the problem statement will be released at the same time.
2. Entries submitted after the contest is closed, will not be considered.
3. Since this is a hiring hack, you are expected to solve the problem on your own.
4. Use of external dataset is not allowed.
5. Participation is free-of-charge.
6. Participant must update their profile details and upload their latest CV.
7. Throughout the hackathon, you are expected to respect fellow hackers and act with high integrity.
8. Slack Live Chat admins hold the right to block any participant found to use foul / disrespectful language.
9. Analytics Vidhya and McKinsey hold the right to disqualify any participant at any stage of competition if participant(s) are deemed to be acting fraudulently.
10. *If due to unforeseen reasons the winner cannot attend or tickets become unavailable, McKinsey will provide a ticket for another analytics conference.

Registration Fee

Free

Feedback ☐

Problem Statement

Your client is an Insurance company and they need your help in building a model to predict the propensity to pay renewal premium and build an incentive plan for its agents to maximise the net revenue (i.e. renewals - incentives given to collect the renewals) collected from the policies post their issuance.

You have information about past transactions from the policy holders along with their demographics. The client has provided aggregated historical transactional data like number of premiums delayed by 3/ 6/ 12 months across all the products, number of premiums paid, customer sourcing channel and customer demographics like age, monthly income and area type.

In addition to the information above, the client has provided the following relationships:

1. Expected effort in hours put in by an agent for incentives provided; and
2. Expected increase in chances of renewal, given the effort from the agent.

Given the information, the client wants you to predict the propensity of renewal collection and create an incentive plan for agents (at policy level) to maximise the net revenues from these policies.

EVALUATION CRITERIA

Your solutions will be evaluated on 2 criteria:

- A. The base probability of receiving a premium on a policy without considering any incentive
- B. The monthly incentives you will provide on each policy to maximize the net revenue

Part A:

The probabilities predicted by the participants would be evaluated using AUC ROC score.

Part B:



The net revenue across all policies will be calculated in the following manner:

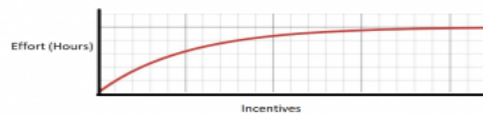
$$Total\ Net\ Revenue = \sum_{Sum\ across\ all\ policies} ((p_{benchmark} + \Delta p) * premium\ on\ policy - Incentives\ on\ policy)$$

(https://s3-ap-south-1.amazonaws.com/av-blog-media/wp-content/uploads/2018/07/Mckinsey_Image.jpg)

- $p_{benchmark}$ is the renewal probability predicted using a benchmark model by the insurance company
- Δp (% Improvement in renewal probability* $p_{benchmark}$) is the improvement in renewal probability calculated from the agent efforts in hours
- 'Premium on policy' is the premium paid by the policy holder for the policy in consideration
- 'Incentive on policy' is the incentive given to the agent for increasing the chance of renewal (estimated by the participant) for each policy

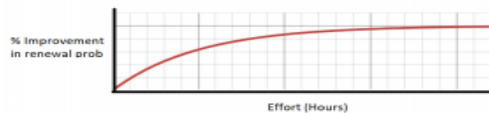
The following curve provide the relationship between extra effort in hours invested by the agent with Incentive to the agent and % improvement in renewal probability vs agent effort in hours.

1. Relationship b/w Extra efforts in hours invested by an agent and Incentive to agent. After a point more incentives does not convert to extra efforts.



Equation for the effort-incentives curve: $Y = 10*(1-\exp(-X/400))$

2. Relationship between % improvement in renewal probability vs Agent effort in hours. The renewal probability cannot be improved beyond a certain level even with more efforts.



Equation for the % improvement in renewal prob vs effort curve: $Y = 20*(1-\exp(-X/5))$

Feedback ☐

Note: The client has used sophisticated psychological research to arrive at these relationships and you can assume them to be true.

Overall Ranking at the leaderboard would be done using the following equation:

$$Combined\ Score = w1*AUC-ROC\ value + w2*(net\ revenue\ collected\ from\ all\ policies)*lambda$$

Where -

$$w1 = 0.7$$

$$w2 = 0.3$$

lambda is a normalizing factor

Public and Private Split:

Public leaderboard is based on 40% of the policies, while private leaderboard will be evaluated on remaining 60% of policies in the test dataset.

Data

train.csv

It contains training data for customers along with renewal premium status (Renewed or Not?)

Variable	Definition
----------	------------



id	Unique ID of the policy
perc_premium_paid_by_cash_credit	Percentage of premium amount paid by cash or credit card
age_in_days	Age in days of policy holder
Income	Monthly Income of policy holder
Count_3-6_months_late	No of premiums late by 3 to 6 months
Count_6-12_months_late	No of premiums late by 6 to 12 months
Count_more_than_12_months_late	No of premiums late by more than 12 months
application_underwriting_score	Underwriting Score of the applicant at the time of application (No applications under the score of 90 are insured)
no_of_premiums_paid	Total premiums paid on time till now
sourcing_channel	Sourcing channel for application
residence_area_type	Area type of Residence (Urban/Rural)
premium	Monthly premium amount
renewal	Policy Renewed? (0 - not renewed, 1 - renewed)

test.csv

Additionally test file contains premium which is required for the optimizing the incentives for each policy in the test set.

Variable	Definition
id	Unique ID of the policy
perc_premium_paid_by_cash_credit	Percentage of premium amount paid by cash or credit card
age_in_days	Age in days of policy holder
Income	Monthly Income of policy holder
Count_3-6_months_late	No of premiums late by 3 to 6 months
Count_6-12_months_late	No of premiums late by 6 to 12 months
Count_more_than_12_months_late	No of premiums late by more than 12 months
application_underwriting_score	Underwriting Score of the applicant at the time of application (No applications under the score of 90 are insured)
no_of_premiums_paid	Total premiums paid on time till now
sourcing_channel	Sourcing channel for application
residence_area_type	Area type of Residence (Urban/Rural)
premium	Monthly premium amount

sample_submission.csv

Please submit as per the given sample submission format only

Variable	Definition
id	Unique ID for the policy
renewal	Predicted Renewal Probability
incentives	Incentives for agent on policy

Feedback ☐

How to make a Submission



How to set a Final Submission on DataHack

[Feedback](#)

McKinsey Analytics Hackathon Rules

1. Entries submitted after the contest is closed will not be considered.
2. Since this is a hiring hack, you are expected to solve the problem on your own.
3. Use of "ID" variable is not allowed
4. Use of external dataset is not allowed

Solution Checker

1. You can use only 10 submissions in a day.
2. Adding comments is mandatory for the use of solution checker
3. Comments will help you to refer to a particular solution at a later point in time.

Final Submission

1. Setting final submission is mandatory. Without a final submission, your entry will not be considered.
2. Code file is mandatory while sending final submission. For GUI based tools, please upload a zip file of snapshots of steps taken by you, else upload code file.
3. The code file uploaded should be pertaining to your final submission.

Data



[📄 Test File \(/contest/mckinsey-analytics-online-hackathon-4/download/test-file\)](#)[📄 Train File \(/contest/mckinsey-analytics-online-hackathon-4/download/train-file\)](#)[📄 Sample Submissions \(/contest/mckinsey-analytics-online-hackathon-4/download/sample-submission\)](#)

Solution Checker

Code File

 No file chosen

Solution File*

(.csv only)

 No file chosen

Solution Description

(max : 180 chars)*

Do you want to show your code on leaderboard?*

Yes ☐ No ☐

You have made **0** submissions out of **10 allowed submissions for the day.**

(HTTPS://TWITTER.COM/SHARE?
UTM_SOURCE=TWITTER_SHARE&UTM_MEDIUM=DH_SOCIAL&URL=HTTP://DATAHACK.ANALYTICSVIDHYA.COM+/I
ANALYTICS-ONLINE-
HACKATHON-4/&TEXT=HEY! I
AM PARTICIPATING IN AN
EXCITING PROBLEM BY
@ANALYTICSVIDHYA
@MCKINSEYANALYTICS
ONLINE HACKATHON
COMPETE AND LEARN.
(DATAHACK)



Analytics Vidhya
Learn everything about analytics



FLASH SALE
Get **55% OFF** on course

**Introduction to
DATA SCIENCE**

Learn basics of Python, Statistics & Machine Learning

SALE ENDS 21ST
JULY, 12:48 PM

Use Coupon Code: **FLASH55**

([https://trainings.analyticsvidhya.com/courses/course-v1:AnalyticsVidhya+DS101+2018T2/about?](https://trainings.analyticsvidhya.com/courses/course-v1:AnalyticsVidhya+DS101+2018T2/about?utm_source=datahackcombannerflash55)
utm_source=datahackcombannerflash55)





ANALYTICS VIDHYA

About Us
(<http://www.analyticsvidhya.com/>)

Our Team
(<https://www.analyticsvidhya.com/about-me/team/>)

Career
(<https://www.analyticsvidhya.com/career-analytics-vidhya/>)

Contact Us
(<https://www.analyticsvidhya.com/contact/>)

DATA SCIENTISTS

Blog
(<https://www.analyticsvidhya.com/blog/>)

Hackathon
(<https://datahack.analyticsvidhya.com/>)

Discussions
(<https://discuss.analyticsvidhya.com/>)

Apply Jobs
(<https://www.analyticsvidhya.com/apply-jobs/>)

Leaderboard
(<https://datahack.analyticsvidhya.com/leaderboard/>)

COMPANIES

Post Jobs
(<https://www.analyticsvidhya.com/post-jobs/>)

Trainings
(<https://www.analyticsvidhya.com/trainings/>)

Hiring Hackathons
(<https://datahack.analyticsvidhya.com/hiring-hackathons/>)

Advertising
(<https://www.analyticsvidhya.com/advertising/>)

Reach Us
(<https://www.analyticsvidhya.com/reach-us/>)

JOIN OUR COMMUNITY :

f
(<http://www.facebook.com/analyticsvidhya>)
47280
(<http://www.facebook.com/analyticsvidhya>)

Followers

(<https://www.linkedin.com/company/analyticsvidhya>)

in
(<https://plus.google.com/+AnalyticsVidhya>)
17125
(<https://plus.google.com/+AnalyticsVidhya>)

Followers

(<https://www.linkedin.com/company/analyticsvidhya>)

Followers

(<https://www.linkedin.com/company/analyticsvidhya>)

Subscribe to emailer

>

© Copyright 2013-2018 Analytics Vidhya.

Privacy Policy (<https://www.analyticsvidhya.com/privacy-policy/>)

Terms of Use (<https://www.analyticsvidhya.com/terms/>)

Refund Policy (<https://www.analyticsvidhya.com/refund-policy/>)

Feedback

⏮

<https://datahack.analyticsvidhya.com/contest/mckinsey-analytics-online-hackathon-4/>

7/7