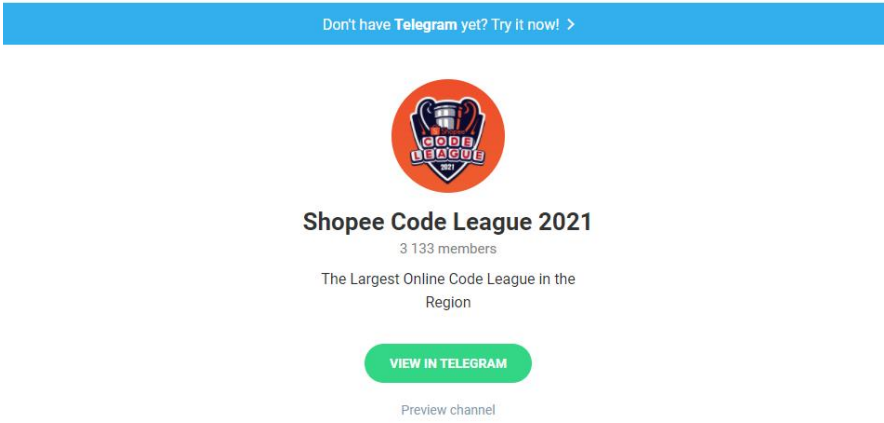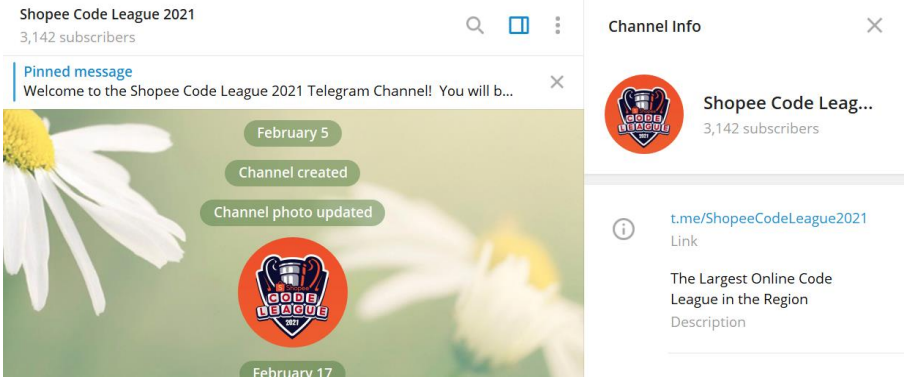**SHOPEE CODE LEAGUE 2021**
KAGGLE COMPETITION GUIDE

Getting Started

To get started, please register on Telegram & Kaggle prior to the start of the competition.
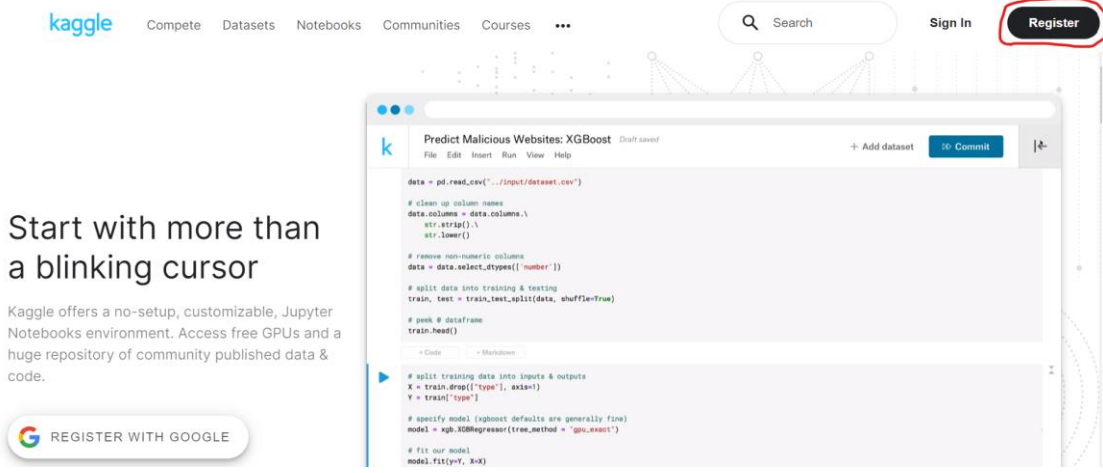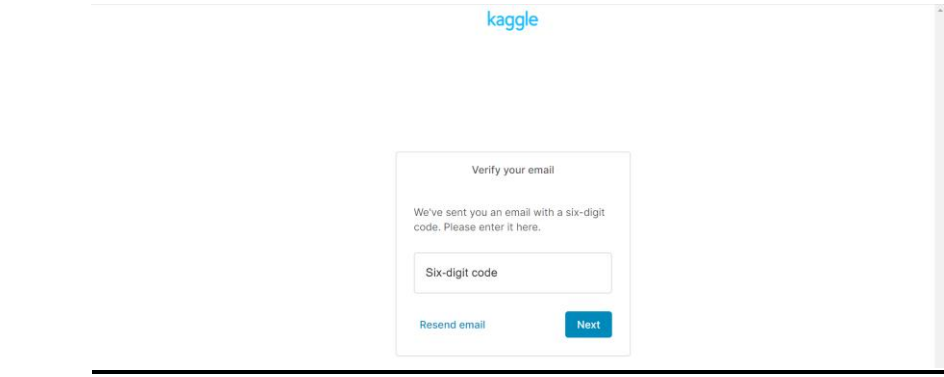
(1) **[Compulsory]** Set up your Telegram Channel to receive important updates and announcements.

Steps	Instructions
1	**Download the app on your mobile** For iOS users: https://apps.apple.com/us/app/telegram-messenger/id686449807 For Android users: https://play.google.com/store/apps/details?id=org.telegram.messenger&hl=en_SG
2	**Enter verification code sent to your mobile**
3	**Access Telegram Web on your Laptop:** https://web.telegram.org/#/login
4	**Join the Shopee Code League 2021 Telegram Channel** https://t.me/ShopeeCodeLeague2021

		Don't have **Telegram** yet? Try it now! ﹥ **Shopee Code League 2021** 3 133 members The Largest Online Code League in the Region **VIEW IN TELEGRAM** Preview channel *Note: The invite link would bring you to this page. Click "Join Channel".
5		**Successful Joining** This should be the page you see upon joining the Telegram Channel.

I. **[Compulsory]** Set up and register for your Individual Kaggle account. Each team member should set up and register for an account. If you already have a pre-existing account, you do not need to register for one again.

Steps	Instructions
1	**Register for a Kaggle Account** Go to Kaggle homepage: https://www.Kaggle.com/ and click register.

2	**Once you've registered, you will see this page. Once you do, check your email for the verification code.**
3	**You may test out the sample Kaggle competition page to familiarise with the features:** https://www.kaggle.com/t/a62cb18d6dde4a48a8c327144049a3d3

Competition Rules

Shopee reserves the right to disqualify any Participants and/or teams at its sole discretion if such Participants and/or teams, where applicable is/are in breach of these Competition Terms and Conditions.

- Participants may use any language they prefer to solve the problem statements - Python, Java, C++, etc
- Participants are allowed to use Internet search engines such as Google during the competition
- Participants are not allowed to enlist remote support from individuals outside of their teams
- For challenges requiring building of models, participants must create and train/finetune their own models. Submissions which directly call any third party APIs on the test set will not be taken into consideration.
- Teams are not allowed to collaborate with each other during the course of the competition, this includes any form of discussion, sharing of notebooks between teams, etc.
- Only submissions made within the time limit for each round will be considered valid, late submissions will not be taken into consideration
- Participants may only compete in the category that they have registered for
- Team names on Kaggle must be the same as the ones used to register for the competition. Only submissions under registered team names will be considered valid

Important: The **top 150** ranking teams based on the private leaderboard must submit their complete scripts after each challenge via email to techsg@shopee.com. Requirements for the submission are detailed in Annex C of the Competition Guide.

- Data Analytics Competition submission, 30 minutes after the end of competition on 6 Mar.
  - [6 Mar] Multi-Channel Contacts

- Data Science Competition submission, by 14 Mar, 11am (GMT+7) / 12pm (GMT+8).
  - [13 Mar] Address Elements Extraction

Note that these rules are non-exhaustive, for more information please refer to the Terms and Conditions on the Competition website.

One account per participant
Each participant shall be limited to only one Kaggle account for the purpose of submission.

No sharing outside of teams

Privately sharing algorithms, codes and/or data for the purposes of the competition, outside of the participant's team is not permitted.

No hand labelling of data

Participants are not allowed to hand label Shopee Data (defined herein) provided to the participants during the competition period. Shopee reserves the right to test the teams' source codes in their respective entries to verify the authenticity of such entry. Shopee reserves the right to revoke the teams' ranking if hand labelling is detected in any entries of such team.

Team size limits

The team size consists of a minimum of 2 to a maximum of 4 Participants per team. Individuals who register for the competition and do not belong to a team shall be allocated to a team at Shopee's sole discretion.

Teams with less than 2 members will be disqualified.

Team mergers

Mergers between teams are not allowed.

External data

Usage of publicly, freely available external data in the teams' entry is permitted.

Eligibility

Participants cannot be an intern, employee, officer, or director and/or independent contractor of Sea Limited (this includes Sea, Shopee, Garena and SeaMoney) during the course of registration and the Shopee Code League until after 20 March 2021.

If you receive an offer from any of the above mentioned entities after registering your team and/or during the course of the Shopee Code League, please write in to techsg@shopee.com to declare.

Entry Requirements

The Competition is open only to registered Participants and Participants can register for the Competition via the Competition Website.

Shopee shall announce the rules and judging criteria of the Competition and any such rules and judging criteria shall be deemed to be incorporated into these Competition Terms and Conditions.

Teams may submit a maximum of five (5) Entries per day for Data Science challenges and a maximum of twenty (20) Entries per day for Data Analytics competitions. Teams can select two (2) final Entries from such teams' past submissions during the Competition Period for judging.

Shopee reserves the final right to reject any Entries in its sole discretion.

Obligations of the Participants

Participant grants Shopee the right to use the Participant's name in connection with such Participant's submitted Entry as well as in connection with any marketing, advertising or promotional material related to the Competition without requiring any further consent or compensation from Shopee.

Participant represents and warrants (where applicable) that it is duly organized, validly existing, and in good standing in the jurisdiction it is formed, that it will comply with all applicable laws in performing under these Competition Terms and Conditions and that it has full capacity, authority and legal right to agree to these Competition Terms and Conditions.

Participants agree that they shall not:

a) use or access datasets provided by Shopee for the purpose of this Competition ("Shopee Data") for any commercial, gambling, or illegal purpose; b) copy, modify, reproduce, license, sublicense, distribute, transmit, republish, exhibit, or display the Shopee Data (except to the extent copying, modifying, reproducing or displaying is necessary and directly related to the Participants' participation in the Competition); c) remove, obscure, or modify any copyright notices or other notices, designations, or attributions within the Shopee Data; or d) use any Shopee Marks or any copyright or other proprietary rights associated with the Shopee Data for any purpose without the express written consent of Shopee "Shopee Marks" means all names, trademarks, logos, symbols, designations, emblems, designs, and trade dress of the Shopee, together with all identifications, labels, insignia or indicia thereof.

The winning team of the Competition shall license the winning Entry and the source code used to generate the Entry under an Open Source Initiative-approved license that in no event limits commercial use of such code or model containing or depending on such code.

Intellectual Property Rights

All intellectual property and proprietary rights in each Participant's Entry shall belong to Participant, save for Shopee's intellectual property embedded in such Entry which shall remain the sole property of Shopee.

Participant grants to Shopee, its and its subsidiaries' and affiliates' agents, contractor and partners, an irrevocable, non-exclusive, royalty-free, worldwide, perpetual and sub-licensable license to reproduce, distribute, transmit, create derivative works of, publicly display and publicly perform the Entry.

Shopee grants to Participants for the Competition Period a non-exclusive, non-transferable and revocable license to use Shopee's Data solely in connection with the Competition. Participants shall not, without the prior written consent of Shopee, alter or modify or create derivative works of the Shopee Data or any of Shopee's intellectual property except as expressedly set forth in these Competition Terms and Conditions.

Nothing in these Competition Terms and Conditions is intended to grant the Participant any rights to use any of Shopee's intellectual property, except as specified in these Competition Terms and Conditions.

Participant warrants that (a) its Entry is their original work and (b) Participant has all rights and ownership or is a licensed user of all algorithms, code and/or data used within the Entry.

Further Information

The following Annexes are enclosed for your information:

- Annex A - Team Formation
- Annex B - Discussion Board Guidelines
- Annex C - Kaggle Guide & Submission Requirements
- Annex D - Common Issues during Submission
- Annex E - Determining your Scores
- Annex F - Common Kaggle Issues
- Annex G - Kaggle Frequently-Asked Questions (FAQs)

TEAM FORMATION

.

1) The team should form their team on the "Team" tab (In the main page of the competition, the tabs are: Overview, Data, Kernels, Discussion, Leaderboard, Rules, Team).

2) Team leader or 1 representative from the team to enter their registered team name in the 'Team Name' field with an added [Category] prefix (e.g. [Student] ShopeeTeam). **Please ensure that the team name on Kaggle is the <u>same</u> as the team name you registered for via the Shopee Code League website. Failure to comply will result in disqualification from the respective Competition.**

Manage Team

Team Name

[Student] Team Coder   Save Team Name

This name will appear on your team's leaderboard position.

3) Under "Merge with other teams", team leader enter the name of your team members. Select the name from the drop-down list and click on 'Request Merge'. (Note: The team leader can only send a Merge Request to one member at a time. Each member must accept the merge request before the leader can add another member to the team)

4) It is recommended that team members choose names that are specific and unique (e.g "Amelia Goh", instead of "Amelia" or "A") to ensure that Team Leaders add the correct members to their teams.

5) **Please ensure that all team members have registered and joined the competition prior to setting up the teams.** Otherwise, the team member's name will not appear in the drop-down list.

Invite Others

⫝ Merge with other teams or invite users to your team by their team name

Amel

Amelia Goh ▲▼

6) Team members accept the invitation from the team leader to form the team by clicking on 'Join this Team'.

Teams Proposing a Merge

You have outstanding merge requests from 1 team:

Amelia Goh
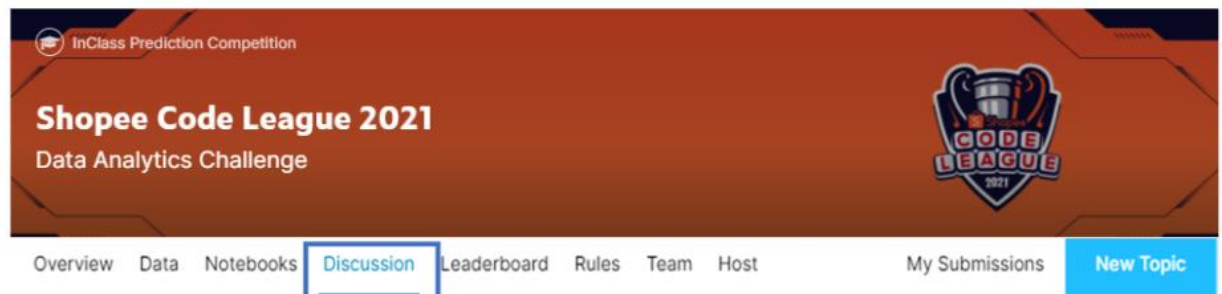led by Amelia Goh   Join this Team

                                                     Reject Invite

7) Team leaders are to check that they have added the right members and team members are to verify the invitation before accepting. Do note that **teams cannot be changed once formed.**

DISCUSSION BOARD GUIDELINES

All technical questions related to the problem statement are to be posted on the Discussion Board. Please DO NOT email techsg@shopee.com on any queries related to the problem statement.

1) Click on the Discussion tab to post problem statement related questions on the Discussion Board.

2) Include the Topic Title and Content in your post
   a) Write a title that summarizes the specific problem, do not use generic titles like "Question" or "I need help".
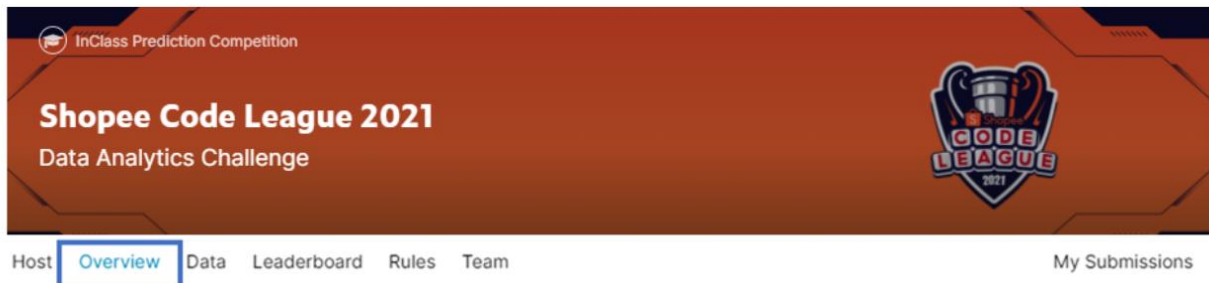   b) See this link for reference on posting a good question

3) Do ensure you have read through the other topics already posted on the page to see if the question has already been answered before posting a new question.
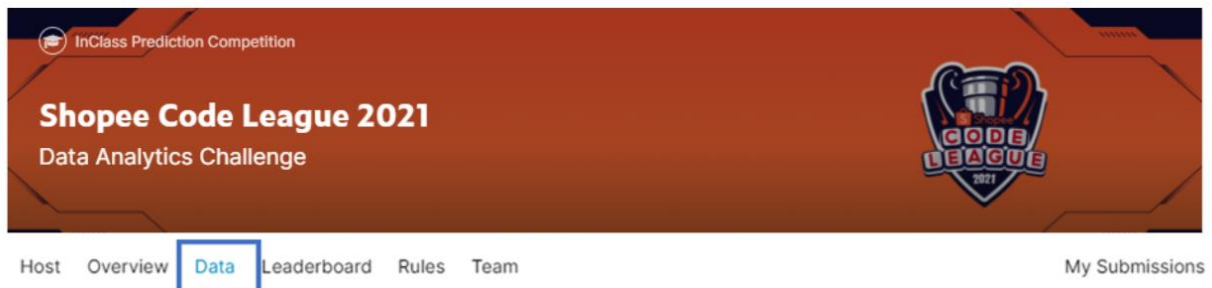
KAGGLE GUIDE & SUBMISSION REQUIREMENTS

You may submit up to **20 times as a team in a day for the Data Analytics problem statement** and **5 times a day for the Data Science problem statement**.

Additional submissions beyond this number will be automatically denied by Kaggle.

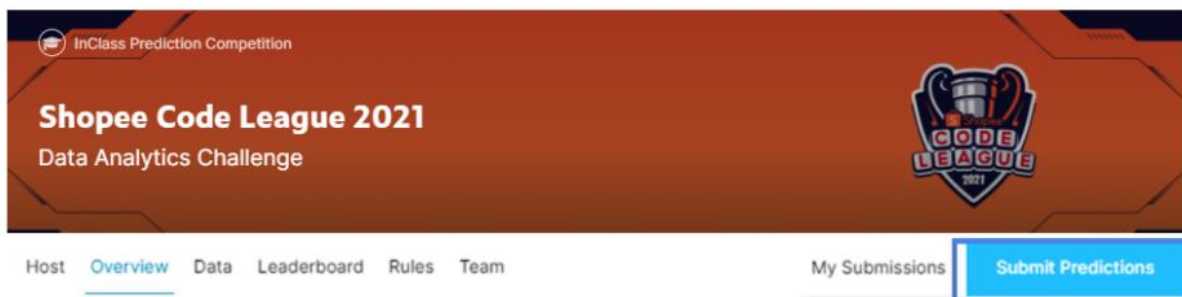1) Read and understand the problem statement and submission requirements in the Overview Tab

2) Download the data required in the Data tab

3) Submit Predictions
Please upload in .csv file on Kaggle.

Step 1: Click on Submit Predictions

Step 2: Upload the required .<mark>csv file</mark> and click 'Make Submission'. **Ensure that your submission meets the requirements before you submit your file.**

Step 1
Upload submission file

File Format
Your submission should be in CSV
format. You can upload this in a
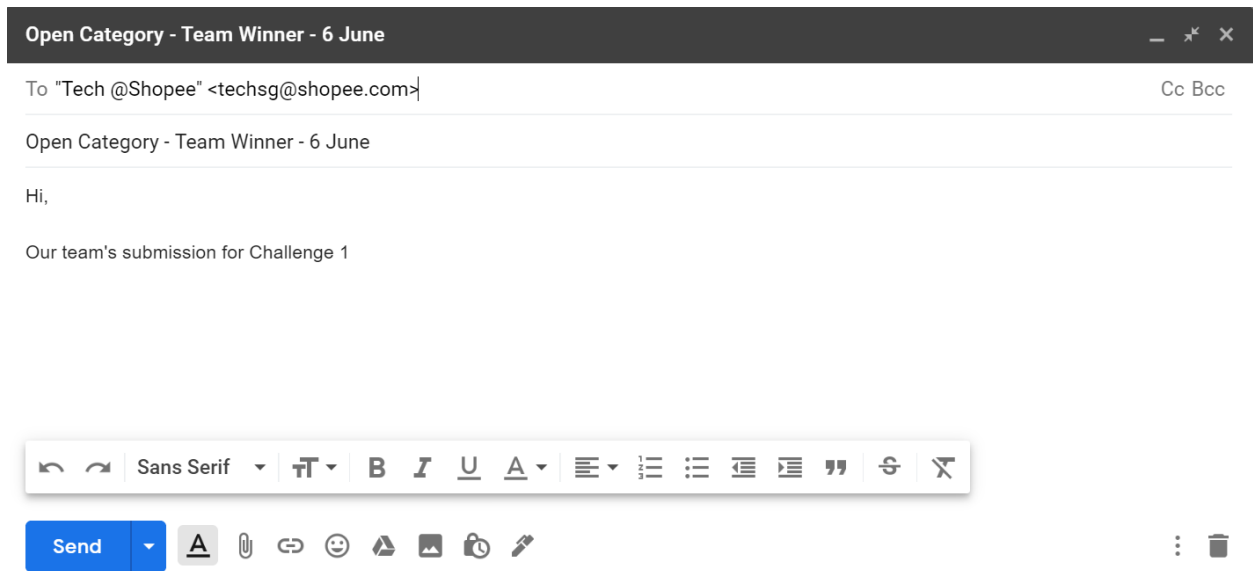zip/gz/rar/7z archive, if you prefer.

Number of Predictions
We expect the solution file to have 3 prediction rows.
This file should have a header row. Please see sample
submission file on the data page.

Note: The **Top 150 teams** based on the Private Leaderboard, regardless of whether you are from Student or Open Category will be required to send in a copy of their script after the end of the competition.

- Data Analytics Competition submission, <u>30 minutes</u> after the end of competition on 6 Mar.
    - [6 Mar] Multi-Channel Contacts

- Data Science Competition submission, by 14 Mar, <u>11am (GMT+7) / 12pm (GMT+8)</u>..
    - [13 Mar] Address Elements Extraction

Any libraries/modules/packages required to run the script must be clearly specified in the script. Email the script to techsg@shopee.com with the subject line: *Category - Team Name (Rank) - Challenge Date* The name of the script should be the same as the subject line. A sample is shown below:

Open Category - Team Winner - 6 June   _ ⟋ ✕

To "Tech @Shopee" <techsg@shopee.com>   Cc Bcc

Open Category - Team Winner - 6 June

Hi,

Our team's submission for Challenge 1

↶ ↷ │ Sans Serif ▾ │ ⊤T ▾ │ **B** *I* U̲ A̲ ▾ │ ☰ ▾ ☷ ☰ ⮜ ⮞ ❞ │ S̶ T̵

Send ▾ │ **A** 📎 🔗 ☺ ◬ 🖼 📋 ✏   ⋮ 🗑

Only one submission per team is required.

Scripts sent have to be **complete** - it should run from start to end without errors and output the required .csv file of predictions for the competition.

COMMON ISSUES DURING SUBMISSION

These are a list of common issues during submission:

1) Wrong Headers

In this example, the required column header was 'new_number'. However, this submission labelled the header as 'newnumber'. Kaggle would show this error when trying to submit a solution with wrongly labelled headers.

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
Wrong Headers.csv	just now	0 seconds	0 seconds	

Failed

ERROR: Column 'newnumber' was not expected (Line 1, Column 4)
ERROR: Required column 'new_number' could not be found

Wrong

index	newnumber
0	1
1	2
2	3

Correct

index	new_number
0	1
1	2
2	3

2) Missing Rows

In this example, the required number of rows is 101, excluding the header row. However this submission only contained 100 rows instead of 101 rows (excluding the header row). Kaggle would show this error when trying to submit a solution with the wrong number of rows.

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
Missing rows.csv	just now	0 seconds	0 seconds	

Failed

Evaluation Exception: Submission must have 101 rows

Wrong

1	index	newnumber
99	97	98
100	98	99
101	99	100

Correct

1	index	newnumber
99	97	98
100	98	99
101	99	100
102	100	101

3) Additional Columns

In this example, the required columns are 'id' and 'new_number'. However this submission contains an additional column labelled 'another_column'. Kaggle would show this error when trying to submit a solution with an additional column.

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
Additional Column.csv	just now	0 seconds	0 seconds	

Failed

ERROR: Column 'another_column' was not expected (Line 1, Column 15)

Wrong

index	newnumber	another_column
0	1	4543
1	2	34323
2	3	2432

Correct

index	newnumber
0	1
1	2
2	3

4) Wrong Index

In this example, the required index (or 'id' column) is a series of numbers from 0-100. However this submission labels the index as 1-100. Kaggle would show this error when trying to submit a solution with the wrong index specified.

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
Wrong Index.csv	just now	0 seconds	0 seconds	

Failed

ERROR: Unable to find the required key value '0' in the 'id' column

Wrong

id	newnumber
1	2
2	3
3	4

Correct

id	newnumber
0	1
1	2
2	3

5) Duplicated Index Value

In this example, all values in the 1st column need to be unique. However, this submission has duplicate values in the 1st column. Kaggle would show this error when trying to submit a solution with any duplicate values in the 1st column.

```
Failed

ERROR: The value '1' in the key column 'stmt_id' has already been defined (Line
3, Column 1)
ERROR: The value '1' in the key column 'stmt_id' has already been defined (Line
4, Column 1)
```

Wrong

stmt_id	ckt_id
1	1234
1	4567
1	7890
4	9012

Correct

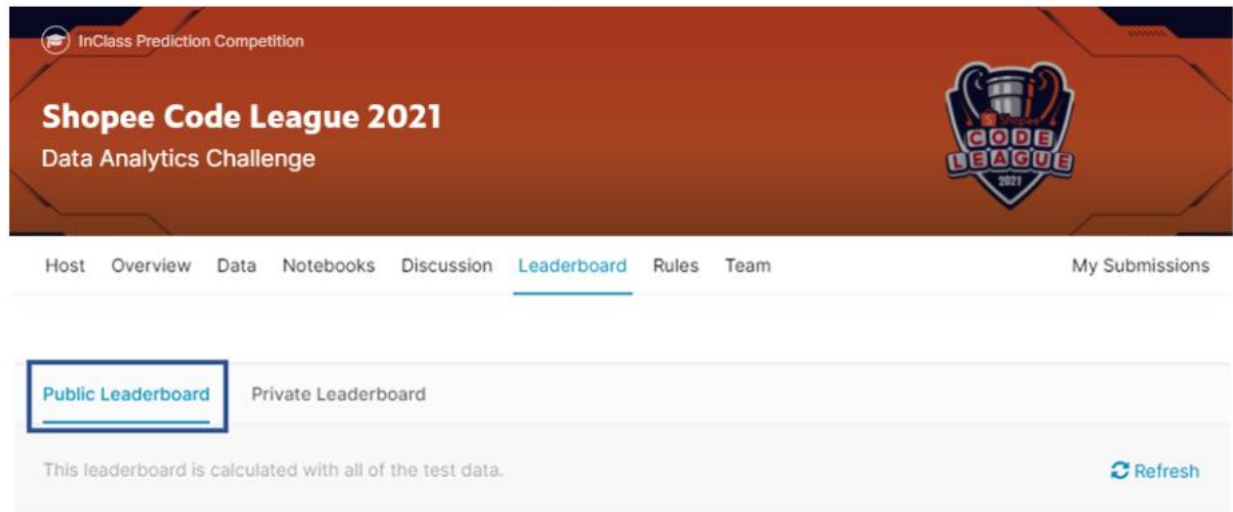| stmt_id	ckt_id
1	1234
2	4567
3	7890
4	9012

Participants are reminded to read the problem overview and submission requirements carefully before attempting the question. No additional time will be given to teams who face problems in submitting their solution.

Important note: While Kaggle provides the Notebook feature for participants to use in order to collaborate on the problems, we **WILL NOT** be providing any support to participants who may face issues in its use. Participants are also reminded it is their responsibility to set their sharing settings on these notebooks to private to prevent others from copying their code.
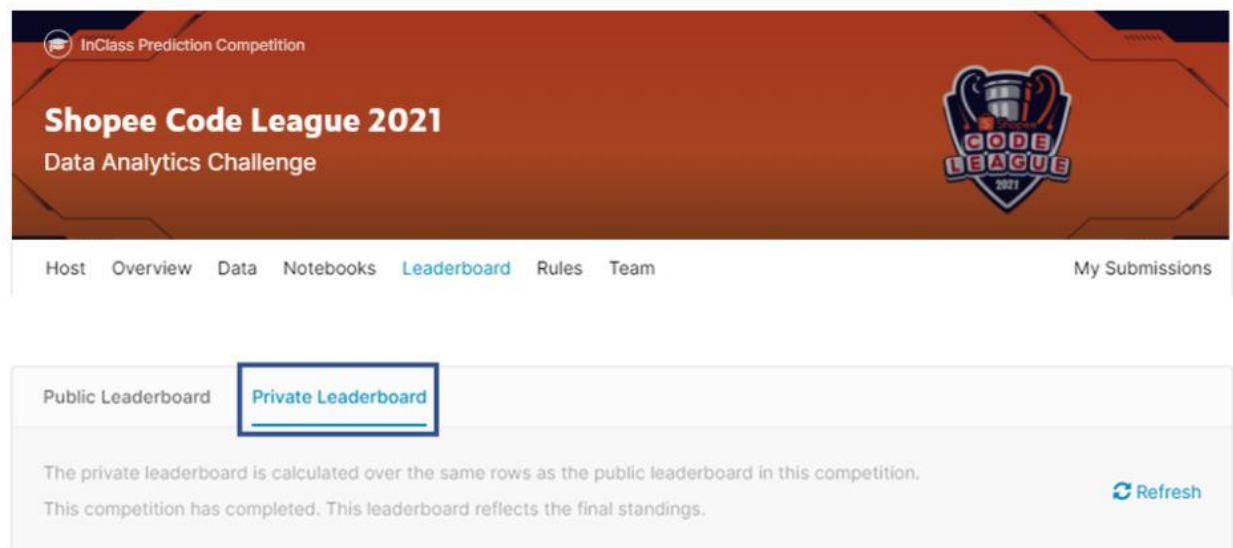
DETERMINING YOUR SCORES

1) While the challenge is active, participants will be able to view their current scores on the **Public Leaderboard.**

   In cases where the test data for the **Public Leaderboard** and **Private Leaderboard** differ, the scores between the two Leaderboards may be different.

2) Once the challenge has ended, overall Rankings for each challenge will be determined by your score under the **Private Leaderboard** page.

3) Regardless of the split between the Public and Private Leaderboards, only the scores on the **Private Leaderboard** will be taken into consideration for calculation of points.

Public Leaderboard   Private Leaderboard

The private leaderboard is calculated with approximately 70% of the test data.
This competition has completed. This leaderboard reflects the final standings.

↻ Refresh

Public Leaderboard   Private Leaderboard

The private leaderboard is calculated over the same rows as the public leaderboard in this competition.
This competition has completed. This leaderboard reflects the final standings.

↻ Refresh

The split of the private and public leaderboard differs for each competition. In general, competitions which involve the creation of models will utilise **approximately 70%** of the test data to prevent overfitting. While competitions which do not require creation of models will utilise **100%** of the test data. Please find the information below:

#	Competitions	Type	Scoring on Leaderboard
1	Multi-Channel Contacts	Data Analytics	Private leaderboard is calculated with **100%** of the test data. Only the Scores on the Kaggle private leaderboard will be taken into consideration for the calculation of points for the Shopee Code League leaderboard.
2	Address Elements Extraction	Data Science	Private leaderboard is calculated with **approximately 70%** of the test data. Only the Scores on the Kaggle private leaderboard will be taken into consideration for the calculation of points for the Shopee Code League leaderboard.

Example

Scores for each competition will be transformed using **Min/Max Normalisation** into points.
Each online competition awards a maximum of 100 points.

In this example, assume that there are only a total of 5 teams in this challenge. The teams have a score of between 0.91577 to 0.91802.

#	Team Name	Notebook	Team Members	Score ❓	Entries	Last
1	HengCherKeng			0.91802	105	6h
2	vladimirsydor[ods.ai][AfterPa...			0.91735	72	20h
3	Как закалялась сталь			0.91655	28	1h
4	MPWARE			0.91580	101	15h
5	[ods.ai] resnet34 is all you need			0.91577	84	4d

We can calculate the Points for each of the teams by applying the following formula:

$$Formula: \left(\frac{value - min}{max - min}\right) x\ 100$$

Teams	Score	Points
HengCherKeng	0.91802	100
vladimirsydor	0.91735	70.222
Kak3akana	0.91655	34.667
MPWARE	0.91580	1.333
[ods.ai]resnet	0.91577	0

COMMON KAGGLE ISSUES

Should you come across any of these issues while attempting the competition, please refer to the guide below on the necessary actions to take.

#	Issue	Solution
1	Accessing the Page ● Too many requests error ● Page not loading	1. Try refreshing the page again at a later time, Kaggle may be experiencing issues with multiple participants trying to enter the page at the same time 2. Check if your teammate is able to access the page, ask your teammate to send you a copy of the problem statement and data
2	Formation of Teams ● Team leader not able to invite team members ● Unexpected token e in json	1. Clear your browser cookies and cache and try again 2. Get a different team member to form the team, at least one person has to set the registered team name. 3. If Steps 1 and 2 do not work out, please email techsg@shopee.com to alert us of this issue with the subject: Team Formation Issue (Your Team Name). a. Include your members details and your current registered team name **BEFORE** the end of the competition
3	Downloading Dataset ● Cannot download dataset	1. Check if your teammate is able to access the page, ask your teammate to send you a copy of the problem statement and data
4	Uploading Solutions ● Scoring takes a long time after making a submission on Kaggle	1. Please be reminded to **make a submission early** in case of technical difficulties. 2. If your team is unable to upload the solutions

		directly on Kaggle, please email techsg@shopee.com with the email subject: Submission Issue (Your Team Name): a. Include your members details and your current registered team name b. Only emails received **BEFORE** the end of the competition deadline will be taken into consideration

KAGGLE FREQUENTLY-ASKED QUESTIONS (FAQs)

This section outlines the Frequently-Asked Questions about the Kaggle Platform. This section has been separated into the following categories:

1. Joining Competition
2. Registration
3. Team Formation
4. Submission
5. Scoring
6. Kaggle Kernels
7. Others

1. Joining Competition

Questions	Answers
How do we attend competitions on Kaggle? Is it automatically assigned?	On the day of each competition, there will be a briefing on zoom 30 minutes before the competition is launched. The competition link will be released real-time on the Zoom chat upon the launch of the competition by our guest-of-honour. Please stay till the end of the opening!
How to join the competition?	The competition link will be released real-time on the Zoom chat upon the launch of the competition by our guest-of-honour. Click on the link to join the competition.
Why do I see "This is a limited participation competition. Only invited users may participate"?	Do use the unique link that will be shared on the Zoom chat or the Telegram channel. When this link has been copied or shared with other participants, the link URL may change.

2. Registration

Questions	Answers
Must the Kaggle account be linked to the email account we signed up with the code league?	For Kaggle, you may use your own individual username to sign up for an account in Kaggle. However, when you are in the competition page, please ensure that you have created the same team name which you have registered for the Shopee Code League. Please check the Team Registration confirmation email for your registered Team Name.
My name in Kaggle was automatically set to a username and not my real name. Does it affect the competition?	
Must the email address for the Kaggle account be the same as the email address used for Code League registration?	
What if I have a Kaggle account under a different email address from the one I used for this Shopee Code League competition? Do I need to sign up for a new Kaggle account?	
Are Kaggle accounts created using a single sign on (e.g. google) usable for the contest?	
Do we need to create new Kaggle accounts if we already have our own accounts?	
Does every team member need to sign up for their own Kaggle account but for Hackerearth only 1 team account is needed, am I right?	Yes for Kaggle, every team member needs to sign up for their own Kaggle account. For Hackerearth only 1 team account is needed. More information on HackerEarth will be released in a separate admin guide.

3. Team Formation

Questions	Answers
How do I join the Kaggle groups? How to make a team in Kaggle?	Once you have clicked 'Join Competition' and accepted the rules. You may click on the Teams tab in order to form teams. Please refer to Annex A of this Kaggle Competition Guide for more details on Team Formation.
How do I pull out of a team, if I was added into the wrong group on Kaggle?	Please check that it is the right team before accepting the merger. You are unable to pull out of the team once you accept the merger.

4. Submission

Questions	Answers
For Kaggle competitions, we are required to send the codes written or only submit the result generated?	The scripts submitted for the competition must be complete - it should run from start to end without errors and output the required **.csv file** of predictions for the competition Any libraries or dependencies should be clearly specified within the script. Please refer to Annex C of this Kaggle Competition Guide for more details.
Can we use tableau, attach screenshot as markdown in Kaggle for data analytics competition?	You may use data visualisations to aid in Exploratory Data Analysis, however they will not be required for submission
How many times can we submit each day?	You may submit up to 20 times as a team in a day for the Data Analytics questions and 5 times a day for the Data Science questions. Data Analytics Questions: [6 Mar] Multi-Channel Contacts

	Data Science Questions: [13 Mar] Address Elements Extraction
Which Kaggle account will be used to count points if each member has an individual account?	All team members should create an account on Kaggle and form teams prior to submission. The highest individual submission score would automatically be chosen as the team's overall score. tr
How will you determine the score for a team on Kaggle if different team members submitted the solutions?	
Do we all need to create a personal Kaggle account or just one team account on Kaggle ?	
How to submit the code in this competition, especially in Kaggle platform?	Once you have joined the competition. You may click the Submit Predictions button to submit your answers. Please refer to Annex C of this Kaggle Competition Guide for more details on submission.
I am unable to make my submission on Kaggle. What do I do?	Please check Annex D of the competition Guide to ensure that you have checked through the common submission issues. Otherwise, do email techsg@shopee.com with the email subject: Submission Issue (Your Team Name) Include your Team Name and members details. Only emails received **BEFORE** the deadline of the competition will be taken into consideration.

5. Scoring

Questions	Answers
What's the rule of calculating team leaderboard score?	Scoring methods on the Private Leaderboard on Kaggle can be found on the Evaluation page for each of the Kaggle challenges.

Questions	Answers
What is the difference between the public and private leaderboard?	For the Data Analytics problem statements, there is no difference between the two leaderboards i.e. the public leaderboard is the same as the private leaderboard. For the Data Science problem statements, the Public Leaderboard is based on ~30% of the dataset while the Private Leaderboard is based on ~70% of the dataset.

6. Kaggle Kernels

Questions	Answers
Is the Kaggle Data Science competition Kaggle kernel only? Or can we use our own machine?	You may run your scripts on your own local machine instead of using the Kaggle kernels.
Will the teams be provided by additional GPU hours on Kaggle during the weeks when the competition is hosted on Kaggle platform?	No. Additional GPU on Kaggle kernels will not be provided during the competition.
Are both data analysis and data science competitions held as in-kernel competitions in Kaggle?	No, you may run your scripts on your local machine. Only the .csv output generated is required for submission on Kaggle. The top 100 teams will need to submit their complete scripts via email to techsg@shopee.com according to guidelines specified in Annex C of this Kaggle Competition Guide.

7. Others

Questions	Answers
I'm new to data science. I've been using Kaggle for a month now, but I never heard about data analytics. Do we need to make a model as well?	Model creation will only be required for the Data Science questions. Data Analytic questions will be focused towards processing of data to find answers to the question posed.
My team member will not have a computer until he gets it repaired - is it ok that he only makes a Kaggle account and joins our team later?	Yes, it is alright since submission is done as a team.