

## 1 Instructions

In the project, you will have the chance to explore an interesting machine learning problem by participating in a Kaggle competition. In this competition, you will predict sales for the thousands of product families sold at Favorita stores located in Ecuador. Specifically, given training data includes dates, store and product information, whether that item was being promoted, as well as the sales numbers, your challenge is to predict the target sales of various products for the 15 days after the last date in the training data. For more details, please refer to official competition website on Kaggle (Store Sales - Time Series Forecasting).<sup>1</sup>

## 2 Teams

- The project should be done in teams comprising of **3-4 members** (preferably 4 members). Note that the competition officially allows teams of up to 5 members, but we are restricting the team size to a maximum of 4 members for all teams from this class.
- Fill out your team information in [this](#) google spreadsheet with your team details. Please log in with your USC email. You can find your team ID in column A within the spreadsheet (Note: please don't change the pre-specified team IDs in the first column).
- Form a team of up to 4 students by **11:59 pm, November 11, 2022** and create your accounts on the official competition website. Make sure that all members of your team are registered under a single team name which begins with `CSCI567_id[TEAM ID]`, for example, `CSCI567_id16`.
- A team can have members from different sections of the class (offline, online, DEN). Please feel free to use Ed to arrange your team.

## 3 Grading

- Your grade is bifurcated as follows: 1) your team's relative rank and score on the leaderboard (40%) and 2) the project report and code (60%). The total base score is 100. The project is worth 20% of the final grade for the class.
- Note that the leaderboard shows the ranking of all teams, not just those in this class. We will only take the relative ranking among the class' teams into consideration.
- Members of the same team will receive the same scores.
- Bonus points in the total score for machine learning class can be earned in the following ways:
  - Early bird bonus: The first 5 teams among all teams in CSCI-567 class at the **HW4 submission deadline** can get 10 bonus points. To get this bonus, at least one of the team members will have to give a 5-minute presentation on their approach in the discussion session on November 17. We will consider the final ranking on the public leaderboard as of **2 pm PST on November 16** for this part of bonus points.

---

<sup>1</sup><https://www.kaggle.com/competitions/store-sales-time-series-forecasting>

- The first 5 teams among all teams in CSCI-567 class in **the final leaderboard** will get 10 bonus points. We will consider the final ranking on the public leaderboard as of **December 10, 2022, 11:59 pm, PT** for grading and this part of bonus points.
- The team that wins the 1st place among all teams in CSCI-567 class will get 10 bonus points.
- If the team wins the 1st place among all the teams on the leaderboard, it will get 10 bonus points.

If a team gets bonus points, all members of the team will be awarded the bonus. bonus policy later.

## 4 Deliverables

- Each team needs to write the project report in NeurIPS format. (6 pages maximum, including references; this page limit is strict). The NeurIPS LATEX format can be found here: <https://neurips.cc/Conferences/2022/PaperInformation/StyleFiles>

Not following the format will lower your grade.

- In your report, you should cover the details of your solutions, including the general ideas, the way of data processing and cleaning, the learning algorithms and models you have tried, the results you obtained, and any other insightful thoughts during the competition. You should also describe how to run your code to get the results.
- All teams are advised to use Python as the programming language for their code. Training should be doable solely on your laptop, but you can use the Kaggle Notebook, [Google Cloud Platform](#), or [Google Colab](#) for running your codes, if needed.

## 5 Deadlines

- We will grade you based on your final released ranking on the online public leaderboard on **December 11**. Please make your final submissions on time. We require teams from this class to submit by **December 10, 2022, 11:59 pm, PT**.
- Each team also needs to submit one PDF copy of the project report and all the code via Gradescope. This needs to be submitted in a single zip file titled `CSCI567_f22_id[TEAM ID].zip` (e.g. `CSCI567_f22_id16.zip`). This is due by **December 14, 2022, 5:00 pm, PT**.

## 6 Policy on collaboration

In line with the rules of the competition, you are only allowed to share code within your own team. Your code will be analyzed to reproduce the results and compare its similarity to code from other teams. Any violation of the USC Integrity and Plagiarism policy will lead to an immediate “F” grade in this class and you might be subject to harsher penalties. However, discussion about approaches between each team members and cross-teams are allowed and we encourage you to actively engage in forums, piazza, and discussion with the Kaggle’s community. This provides a great learning opportunity for you to explore how to collaborate and research new approaches to tackle hard challenges.

Using resources from the internet is acceptable, but you must cite and acknowledge any resources you used, and should understand and analyze everything you use as part of your report.