Team name: Team Capybara
Team members: Yu Chen, Zifei Dong, Ning Pan, Sifan Tao, Yao Yao
Date:

This journal has to be uploaded to 1) a **dropbox folder** (https://www.dropbox.com/request/oMzqzRNvauGRuF97kjyh). **AND** 2) posted to your project website every two weeks by 11:59 pm on the second Friday (except the spring break). When uploading to the dropbox folder, change the file name to "Team_Capybara _MM-DD-2023" where MM = month and DD= day of upload).

Team roles for this report (write down name):

Facilitator(s): Zifei Dong

Recorder(s): Yu Chen

Deliverer(s): Ning Pan, Yao Yao

Planner(s): Sifan Tao

See last page for description of roles. Obviously one person can take more than one role or there can be more than one person per role or make your own roles!

0. **Describe briefly what the main goal of your team is (so the peer reviewer has some context). E.g. we are working on image classification for blah de blah. Our goal is blah de blah etc. In the initial part of the semester before your proposal it is ok to put down "we are still coming up with ideas on team project".**

   We are working on a prediction model that accurately predicts the sale price of a house based on various features. This model can be used by various stakeholders such as real estate agents, homeowners, and property investors to make informed decisions about buying, selling, or investing in a property. The ultimate goal of our project is to create a model that performs well on new and unseen data, is interpretable, and provides insights into the most important factors that drive house prices. This will enable stakeholders to gain a better understanding of the housing market and make informed decisions based on the insights provided by the model. We expect our model to have the

potential to provide valuable insights and assist stakeholders in making informed decisions related to the housing market. Meanwhile, Yu Chen will use an image dataset to do classification as a side project for the one above, if this goes well, it might be also on the final presentation.

I.   **What was done during the report period regarding the project:** If you want to include code include this in the Appendix. Describe what the group did (including contributions of individual team members) with regards to the group project during this report period. Give enough details so I understand what you folks have been doing over the week. Include dates of your meeting(s) and who met on these days.

Meet date: March 5: all members, March 7: Zifei Dong, Sifan Tao, Yao Yao, March 9: all members.

Zifei Dong: Found the data. Composed and refined EDA RMD, project proposal, and biweekly report. Build GitHub Page for Html, Word, and Pdf hosting. Reform cloud storage and database. Add the "Inspecting relationships over a subset of the observations" section to EDA.

Sifan Tao: Researched techniques we would like to apply in our project. Looked at different methods for building and testing models. Wrote preliminary ideas on techniques in our project proposal. Completed the visualization part of our EDA.

Yao Yao: Summarized data source and data description part. Wrote the goals and motivation in our project proposal. Completed the distribution analysis part of our EDA.

Ning Pan: I have investigated various predictors that may have an influence on house prices. Furthermore, I cleaned up the dataset and performed correlation analysis as part of the exploratory data analysis to identify any significant relationships between the predictors and the house prices.

Yu Chen: Browsed different image datasets, currently interested in the wiki art data.

II.   **What were obstacles faced if any in working on the project?** This could be technical (like not being able to implement or understand particular techniques) or time issues (midterms for other courses etc).

Yu Chen had three exams this week, the original proposal discussed on Sunday was not achievable due to the complexity of the data, so we changed the proposal to predicting housing price.

III. **What is the plan for the next reporting period including what each team member is planning to work on. Describe goals and potential timelines (" I plan to finish understanding x to see if it can be implemented for our project by Wednesday etc".)**

Zifei Dong: Reading and imitating model training online. Explore different techniques on the house model and may consider selecting the most influenced variables within the spring break.

Yao Yao: I plan to finish researching techniques for predicting house pricing models. Set up our project framework together in the process of research. Plan to meet after Spring break and set up times for group members to meet to facilitate.

Sifan Tao: Research on how to combine and compare the results from different regressions. Learn more about model stacking regarding using predicted results to remodel again.

Ning Pan: My plan is to explore various models that can be used to predict house prices. Additionally, I aim to enhance the quality of the dataset by performing data cleaning.

Yu Chen: Learn SOTA models for image classification, and how to implement those. Resize the image dataset for training and run a basic model during spring break. For the housing data, I will explore different regression methods.

While in the biweekly document above you will describe what your team did with regards to the team project (with proper attributions of who did what in the week) there are 4 pre-defined roles. I urge you to have different people do these jobs every week so that you gain experience in each of the jobs. There can also be more than one person per job for example 2 people recording the weekly journal.

Facilitator: Manages the group for this week including setting up times for group members to meet, making sure everyone has a say in the meetings etc.

Recorder: Person in charge of recording the meetings as well as the happenings of the past two weeks and describing what was accomplished in the meeting and writing up this report.

Deliverer: Person in charge of checking the entire report and uploading the file to dropbox folder and project website, as well as the representative of the group getting in touch with the instructor.

Planner: Person in charge of what will be happening next two weeks as well as thinking about longer term goals (what more needs to be done for the project).

Team contact: Person I can email if I see any issues in the biweekly report instead of mass spamming everyone in the team.