

# 

# **KPI Prediction Tool: A Look Into The Future**

# **Progress Report #1**

## **Economics 599/611**

## **Department of Economics**

## **University of Calgary**

**CHANGES SINCE PROPOSAL**

The noticeable changes made since the submission of the KPI Project Plan are as follows:

1. Following mid-February, Team 4 stopped having regular meetings outside of class time as everyone was completing individual tasks, instead, members of the team were expected to provide regular updates on Slack.
2. Beginning the reading break, February 21st/2022, the Gantt chart was updated to provide more time for the workshop teams, which is another group project in the Economics 599.88/611 course that encompassed different team members.
3. Dr. Crost extended the deadline for the first Progress Report from March 10th/2022, as indicated in our Project Plan’s Milestones and Deliverables, to Thursday, March 17th/2022.
4. Although a decision has not been made regarding the evaluation method, Critical Mass is no longer evaluating our project against a benchmark model.

**PROGRESS IN EACH MILESTONE**

Since the start of the project, we have maintained good communication with the representatives of CM: we continue to give regular –weekly or bi-weekly– project updates, address pressing issues encountered during the project, and if needed provide clarification regarding our project to CM. Overall, we have managed to complete tasks as they were set in the deadline however, Team 4 did face its challenges such as the ongoing workshop projects diverted the team's attention away from the KPI project. This was especially true for members of our team who had to present within the first two weeks of returning to school from reading break.

During Phase 2 and Phase 3 following the submission of the KPI Project Plan the data received from CM was cleaned and prepared for modeling through the process of data wrangling and transformation. Currently, about 60% of our Exploratory Data Analysis (EDA) has been completed, with the remaining EDA intended to be done with a pandas function. However, Team 4 encountered a roadblock in the data wrangling phase. The issue was determined to be the format in which our target variables (KPIs) were displayed. For example, many of our features as well as the KPI’s are recorded weekly over the course of two years. Additionally, not all creatives launch at the same time, making the data wrangling process more difficult. To fix this issue, we normalized all date variables to begin in the same week, week 1, lasting up to week 48, the longest time a creative was in the market. Thereby, allowing us to analyze our data on common terms. We are now ready to proceed to Project Implementation, particularly focusing on forecasting model, model evaluation, user testing, and GUI, and lastly, work on our Progress Report 2.

The following paragraphs explain the specific tasks completed and those that have yet to be completed throughout Phase 1 to Phase 4 as outlined in our Gantt chart. (Appendix A)

**Phase 1**

Phase 1 spanned the month of January and consisted of project initiation and planning, including the submission and getting approval for the Partner Agreement Form and Project Proposal. Phase 1 is complete

**Phase 2**

Phase 2 spanned the month of February and consisted of understanding the data. After cleaning the data and some feature engineering, we did some mechanical EDA. To further explore the data, we have found a useful function in pandas that will automatically analyze our features. Currently, we are running an EDA profile report using a pandas module named pandas. profiling. This module uses the function ProfileReport(df) to generate a report including critical values and distributions of features, correlations between features, outliers/anomalies, and more. The goal is for this report to help us evaluate feature importance and provide some ideas for feature interaction. Although we will be progressing to Phase 3, we will revisit Phase 2 simultaneously as we test our predictive model.

**Phase 3**

Phase 3 spans the month of March and consists of model selection, evaluation, tuning, project implementation, and Progress Report 2. Although both feature engineering and feature importance are partially complete, we are relying on the profile report to tell us more and get us closer to understanding feature importance. Also, we will simultaneously create and remove features as we test and improve our machine learning (ML) model over the following weeks.

**Phase 4**

Finally, Phase 4 will span over the remainder of the semester –the first two weeks of April –. This phase primarily consists of the creation of our final report, validating our code externally, and finally, creating and then presenting our KPI project to our classmates, professor, and project partner, CM.

**PROTOTYPE APP**

We will provide an update in Progress Report 2.

**MODEL PSEUDOCODE**

We have not begun developing our model code, however, we can detail code a potential model candidate could follow.

1. Split data into training, validating, and testing groups
2. Instantiate desired model
3. Loop through k-fold cross validation combinations of training and validating sets
   1. Fit model on training sets
   2. Predict target variables on the validation set
   3. Save error/accuracy metric
4. Compute the mean error/accuracy overall folds

* Assuming this is our final experimental model, we select the best-performing one with the lowest/highest error/accuracy.

1. Train best model on training and validating sets combined
2. Predict target variables on the testing set
3. Generate error/accuracy metric

* Splitting the data into three groups instead of two will allow us to be more confident in the performance of our final model.

**APPENDICES**

**Appendix A**

| **WBS NUMBER** | **TASK TITLE** | **TASK OWNER** | **START DATE** | **DUE DATE** | **PCT OF TASK COMPLETE** |
| --- | --- | --- | --- | --- | --- |
|
|
| **1** | **Project Initiation and Planning** |  |  |  |  |
| 1.1 | Group creation | All | 12/01/22 | 13/01/22 | 100% |
| 1.2 | Group meeting | All | 15/01/22 | 15/01/22 | Ongoing |
| 1.3 | Meeting with project partner | S, M, F, W | 20/01/22 | 20/02/22 | Ongoing |
| 1.4 | Outline tasks and assign roles | S, M, F, W | 24/01/22 | 27/01/22 | 100% |
| 1.5 | Gantt Chart | W, M, F | 24/01/22 | 29/01/22 | 100% |
| 1.6 | PPT Presentation | S, M, F, W | 29/01/22 | 10/02/22 | 100% |
| 1.7 | **Project Proposal** | All | 22/01/22 | 08/02/22 | 95% |
| **2** | **Understanding Data (Progress Report 1)** |  |  |  |  |
| 2.1 | Data Wrangling | Y, S, W, M | 08/01/22 | 12/02/22 | 100% |
| 2.2 | Data Cleaning I | Y, S, W, M | 08/02/22 | 15/02/22 | 100% |
| 2.3 | Data Analysis | S, W, M | 08/02/22 | 17/03/22 | 60% |
| 2.4 | **Progress Report 1** | F, S, W | 09/02/22 | 17/03/22 | 100% |
| **3** | **Project Implementation (Progress Report 2)** |  |  |  |  |
| 3.1 | Data Cleaning II | Y, W | 01/03/22 | 25/03/22 | 50% |
| 3.2 | Data Preparation | S, F | 14/03/22 | 21/03/22 | 50% |
| 3.3 | Forecasting Model | M, W | 16/03/22 | 30/03/22 | 10% |
| 3.4 | Model Evaluation | S, M | 16/03/22 | 31/03/22 | 0% |
| 3.5 | User testing and GUI | M | 24/03/22 | 04/04/22 | 0% |
| 3.6 | **Progress Report 2** | F | 14/02/22 | 31/03/22 | 0% |
| **4** | **Final Code, Report, and Presentation** |  |  |  |  |
| 4.1 | Modify project based on feedback |  | 18/03/22 | 04/04/22 | 0% |
| 4.2 | **Report + Presentation** |  | 19/03/22 | 11/04/22 | 0% |
| 4.3 | **Code + User-guide** |  | 20/03/22 | 14/04/22 | 0% |