

# 

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

# **Progress Report #2**

## **Economics 599/611**

## **Department of Economics**

## **University of Calgary**

**CHANGES SINCE PROGRESS REPORT 1**

The noticeable changes made since the submission of Progress Report 1 are as follows:

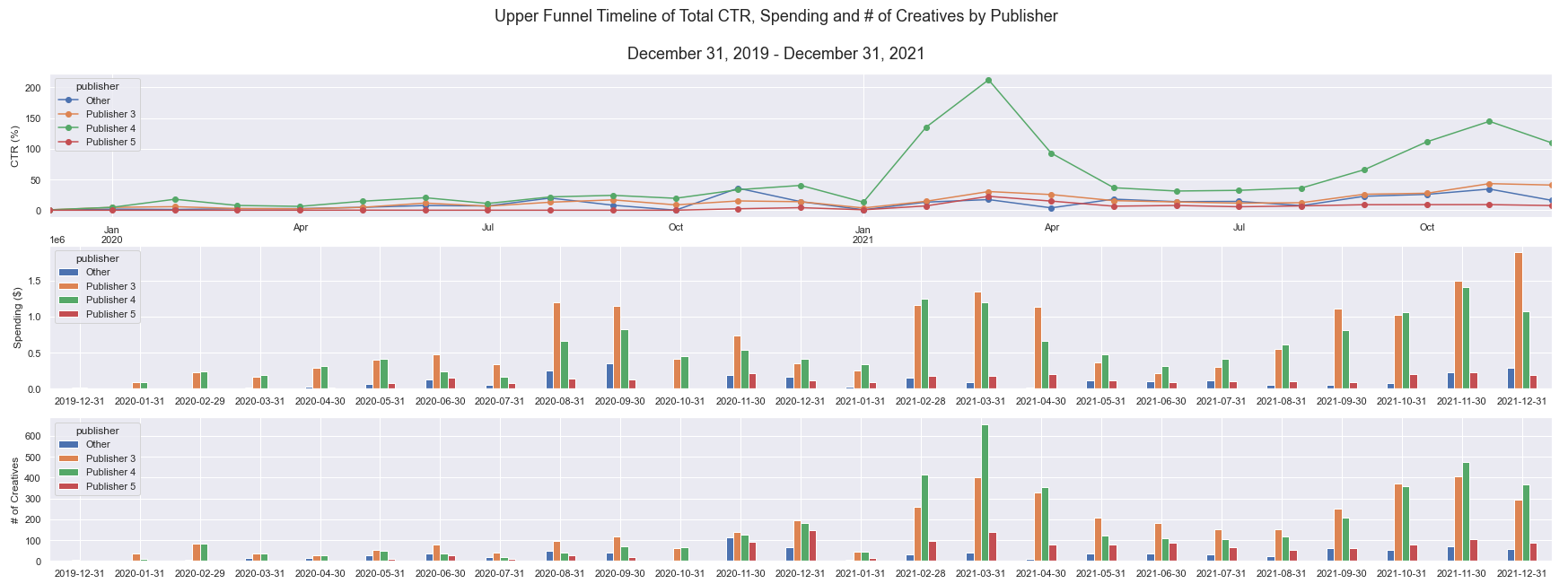
1. The use of the Gantt Chart and regular team meetings decreased as Team 4 members were aware of their responsibilities and were working individually.
2. Similar to Progress Report 1, Dr. Crost extended the deadline for Progress Report 2 from March 31st/2022, as indicated in the Project Plan’s Milestones and Deliverables, to Thursday, April 7th/2022.
3. Due to runtime issues with the raw data, CM suggested we set aside the Display channel and use the Social channel only. This reduced our number of observations by almost 11,000 leaving us with approximately 2,500 observations.
4. Working on selecting a model however, for now we have started working with a Linear Regression model.
5. An Exploratory Data Analysis (EDA) analysis was carried out and currently, we are working to finalize the three primary learning objectives.
6. We have started off our prediction tool by building three linear regression models, one for each funnel.

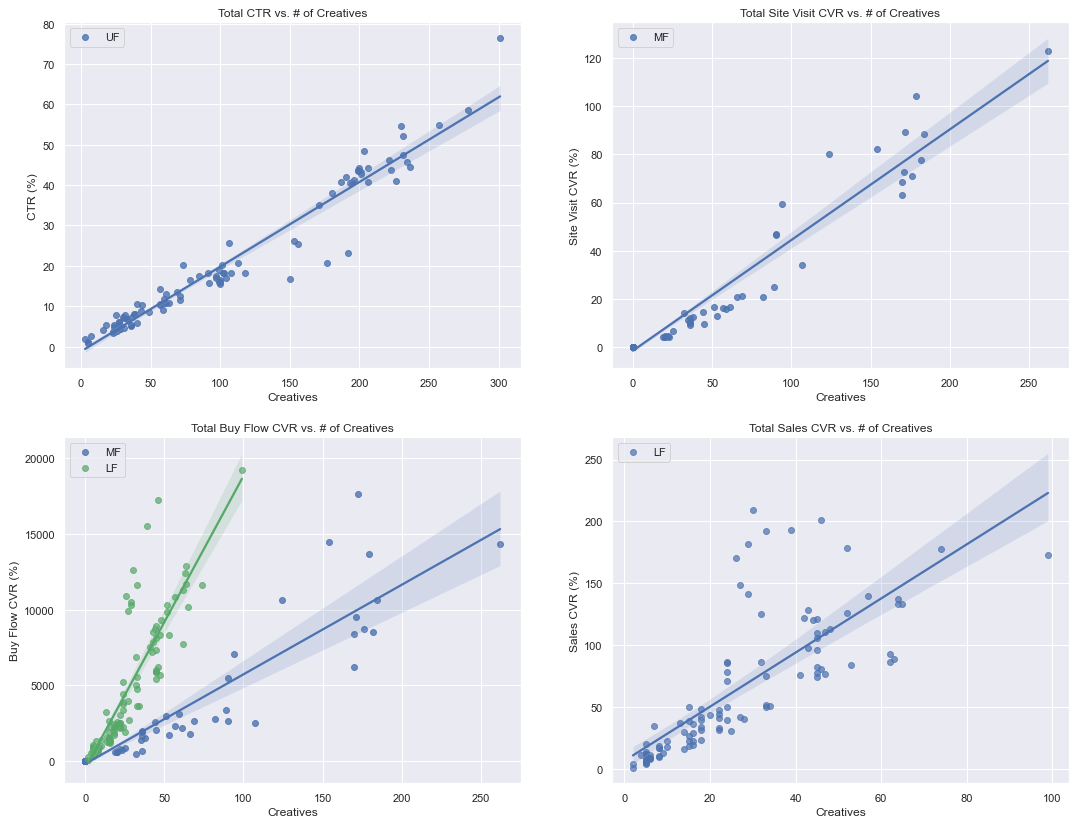
**PROGRESS**

**Phase 3**

Throughout the completion of the project, Team 4 maintained good communication with Critical Mass (CM). A decision was also made following the submission of Progress Report 1 to maintain the regular bi-weekly meeting with CM which allowed Team 4 to provide project updates, address pressing issues encountered during the completion of the project, and if needed provide clarification regarding the KPI project. Overall, we are progressing according to the schedule set out in the Gantt Chart. However, Team 4 did face a drawback because of an unexpected family emergency that affected the completion of the EDA.

So far, Team 4 has been working on completing Phase 3 which entailed cleaning and preparing the data; forecasting and evaluating a chosen model; working on Progress Report 2; and getting started on the User testing and GUI. Currently, the entire EDA has been completed using python. Also, as CM requested, Team 4 answered three primary learning objectives regarding the data. The first one being the most important features and their impact towards KPI prediction. Although there are some preliminary results, they are currently not up to our standards to disclose. The second objective was to find the optimal number of creatives per funnel and publisher. Below is a graph depicting the behaviour of upper funnel Click Through Rates (CTR) by publisher when advertisement spending and number of creatives changes.

We can see that creatives in publisher 4 have much higher CTRs than creatives in publisher 3, and publisher 3 creatives have little to no impact on CTR. Team 4 is working to determine whether spending has an effect on the KPI of publisher 4. Since it follows the same trend as the number of creatives, it does not appear to influence CTR of publisher 3. The third objective was to find the impact of the number of creatives in the market on performance. To answer this question, we ran funnel specific regressions of KPI on the number of creatives in the market while keeping in mind time series regression assumptions. We found that, on average, adding one more upper funnel creative relates to a total weekly CTR increase of 0.1739 points. Next, we plan to change this from total weekly to average weekly as CM representatives recommended. 



Baseline Linear Regression Model

We have started off our prediction tool by building three linear regression models, one for each funnel. The first linear regression models were built to predict peak KPIs by regressing on all covariates.

* In the Upper Funnel the KPI of interest is CTR’. This regression for peak CTR left us with an R-squared coefficient of approximately 0.16
* In the Mid Funnel the KPI of interest is ‘Site Visit Conversion Rate (CVRSV)’. This regression for peak CVRSV left us with an R-squared of approximately 0.20
* In the Lower Funnel the KPI of interest is ‘Sales Conversion Rate (CVRS)’. This regression for peak CVRS left us with an R-squared of approximately 0.26
* Another KPI of interest is ‘Buy Flow Conversion Rate (CVRBF)’, which is both a Mid & Lower Funnel KPI. This regression for peak CVRBF left us with an R-squared of approximately 0.42

It is clear that a linear regression model is not currently giving us enough predictive power. We are moving forward to additional models to improve the funnel specific predictions. However, it is noteworthy to see that as we move down the funnel we are getting more predictive power using a linear model.

**Phase 4**

Despite starting on the coding component of the project in early March, Phase 4 had a delayed start. Following the submission of this report, Team 4 will be working on validating the code externally, developing the GUI, and finalising the final report and Power-point presentation, which will be later presented to CM and Dr. Crost. Team 4 will also work on selecting a model, which will require some pre-processing steps and testing.

**PROTOTYPE APP**

If time permits, Team 4 will try to use Tableau after building the model. This will require us to make a local server, then upload the model to the Tableau before tweaking the model, and graphing the performance over time to provide visuals. However, if constrained by time, Team 4 will instead create a wireframe of what the dashboard will look like in Tableau.

**MODEL PSEUDOCODE**

1. Split data into training, validating, and testing groups
2. Loop through k-fold cross validation combinations on the training sets only
   1. Create a pipeline with sklearn’s <make\_pipeline> function and pass in as arguments
      1. Preprocessing methods to data
      2. Desired model to fit on preprocessed data
   2. Predict target variables on the validation set of the training
   3. Generate error/accuracy metric
3. Compute the mean error/accuracy over all folds

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

1. Predict target variables on validation sets
   1. If good model, continue
   2. If bad model, repeat steps with a new approach
2. Using training and Validating datasets repeat steps 2 and 3
3. Predict target variables on the testing set
4. Generate error/accuracy metric

*Note:*

*The only thing new here is we are cross validating on a* ***process*** *rather than a model. We can also take it a step further and perform hyperparameter tuning on pre-processing steps. Thus, giving us the ability to create k-fold machine learning processes and average their performance.*

**APPENDICES**

**Appendix A**

| **WBS NUMBER** | **TASK TITLE** | **TASK OWNER** | **START DATE** | **DUE DATE** | **PCT OF TASK COMPLETE** |
| --- | --- | --- | --- | --- | --- |
|
|
| **1** | **Project Initiation and Planning** |  |  |  |  |
| 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 |
| **3** | **Project Implementation (Progress Report 2)** |  |  |  |  |
| 3.1 | Data Cleaning II | Y, W | 01/03/22 | 25/04/22 | 100% |
| 3.2 | Data Preparation | S, F | 14/03/22 | 21/03/22 | 100% |
| 3.3 | Forecasting Model | M, W | 16/03/22 | 14/04/22 | 10% |
| 3.4 | Model Evaluation | S, M | 16/03/22 | 14/04/22 | 0% |
| 3.5 | User testing and GUI | M | 24/03/22 | 14/04/22 | 0% |
| 3.6 | **Progress Report 2** | F | 14/02/22 | 07/04/22 | 100% |
| **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 | 14/04/22 | 0% |
| 4.3 | **Code + User-guide** |  | 20/03/22 | 14/04/22 | 25% |