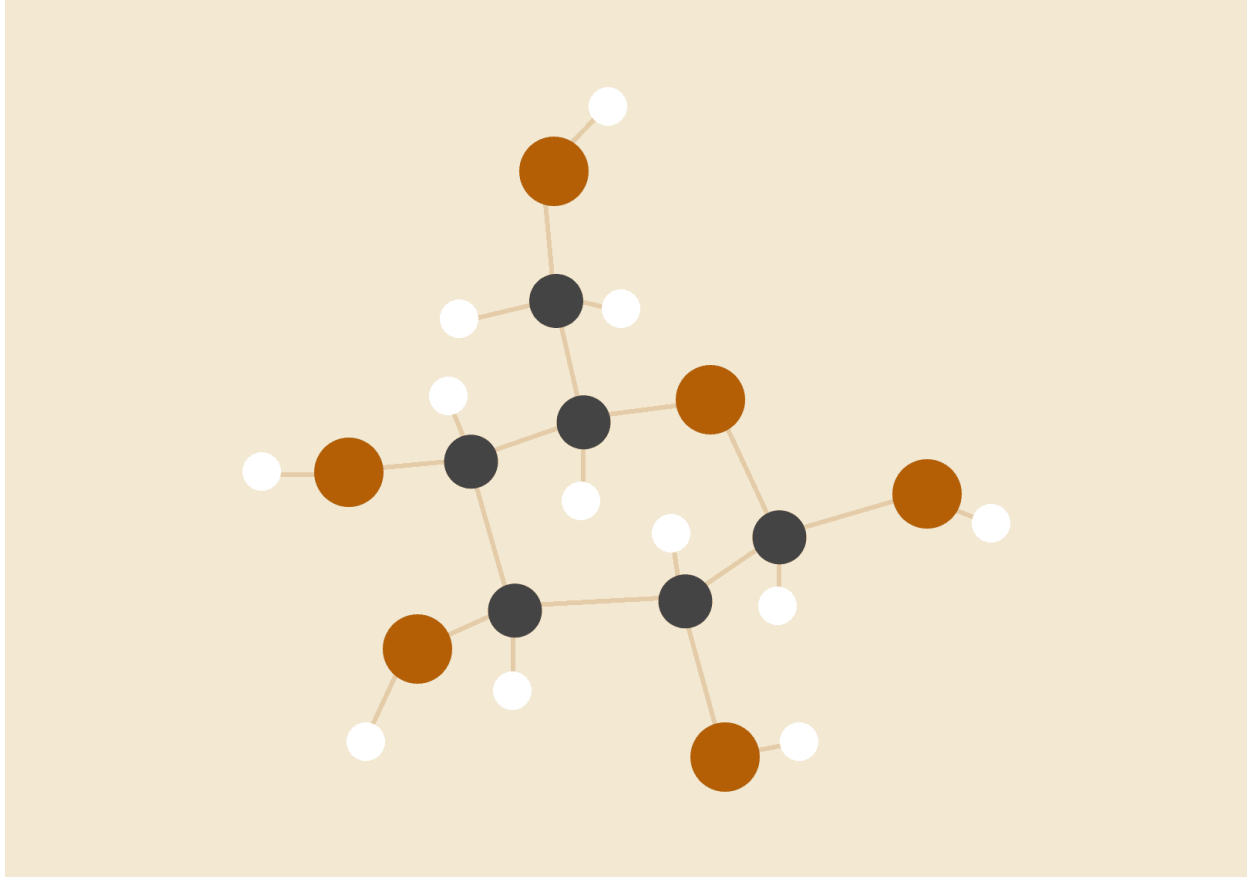


GROUP MINUTES



BY TEAM 23:

- **STELLA MUKUHI**
- **DANIEL IFEDIBA**
- **UBASINACHI ELEONU**
- **JEFF OUMA**
- **OBINNA UZOEGBU**
- **GEORGE KIBE**

MEETING 1

THE MEETING HELD ON MONDAY, 22ND AUGUST 2022, AT 12:00 MIDDAY SOUTH AFRICAN TIME

Matters discussed:

1. Members introduced themselves
2. The task at hand was thoroughly discussed as well as the project overview
3. The dataset was also discussed to understand the features and target variable
4. It was agreed that a Github repo should be created
5. Every member would create a personal Github branch off the main branch
6. Roles were assigned to each team member
7. Members agreed on times to meet during the week
8. Organization of the Trello board was discussed
9. The first base notebook would be created and pushed to the GitHub repo.

MEETING 2

MEETING HELD ON WEDNESDAY, 24TH AUGUST 2022, AT 8:00 PM SOUTH AFRICAN TIME

Matters discussed:

1. Daniel created the first base notebook and pushed it to the GitHub repo
2. Daniel took through the base notebook and the RMSE & R2 scores
3. Other members agreed to continue working on the base notebook to try to improve the performance metrics
 - Stella, Obinna, and George to further work on data cleaning and preparation
 - Ubasinachi to work on the EDA section of the notebook
 - Jeff Ouma to work on the model training and evaluation to better improve the performance metrics
4. Everyone else to push their contributions to their branches on the GitHub repo and write a clear commit message against them.

MEETING 3

MEETING HELD ON FRIDAY, 26TH AUGUST 2022, AT 8:00 PM SOUTH AFRICAN TIME

Matters discussed:

1. Meeting was held to discuss and further elaborate on the meaning of each feature in the dataset and its significance in improving the machine learning models' accuracies
2. Some features were the best thought of as irrelevant and, thus, should be dropped

MEETING 4

MEETING HELD ON MONDAY, 29TH AUGUST 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Meeting was held to further discuss how to improve the accuracy of the model, using several methods, e.g. more data cleaning, feature selection, and data engineering

MEETING 5

Meeting held on Wednesday, 31st August 2022, at 10:30 am South African Time

Matters discussed:

1. Discussion centered around how to do feature engineering for the investment variables. The investment variables were in a percentage format, so it was agreed upon that the investment variables would be converted to their original format, to help increase the accuracy score

MEETING 6

MEETING HELD ON FRIDAY, 2ND SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. The decision to convert the investment variables from the percentage format to their original format was wise since the R2 score increased from 0.91 to 0.99
2. Nomfundo informed us that we need to prepare for a presentation, where we will formally present our ideas and progress to an external supervisor. Preparations are now in progress

MEETING 7

MEETING HELD ON MONDAY, 5TH SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Preparations are now in progress on how to conduct the presentations, which are scheduled for Friday, the 9th of September. Key items to include in the presentation include:

- i. Brief introduction to the project
- ii. Elaboration of the progress we have made so far and the findings we have achieved
- iii. Discussion on the models we have used to perform the machine learning
- iv. Presentation of the performance metric scores we have obtained so far,
i.e. R2 and RMSE scores
- v. Outline how we seek to improve our scores

MEETING 8

MEETING HELD ON WEDNESDAY, 7TH SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Preparations are still in progress on how to conduct the presentations, which are scheduled for Friday, the 9th of September.
2. We also agreed to use the Power BI platform to conduct our presentation. A brief run of this presentation will be done a day prior (on Thursday 8th September 2022) with Nomfundo, so that she can our work beforehand and point us in the right direction

MEETING 9

MEETING HELD ON FRIDAY, 9TH SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. We held a discussion with Nomfundo Manyisa regarding the first presentation we did to the external supervisors, i.e. Willie and Paul
2. Presentation seemed one-sided and improvements should be done to ensure full engagement with all members of the team
3. It was agreed that the presentation will be done using canvas slides (instead of Power BI) to ensure all members have access to the slides and can add/ edit them
4. Canva slides will also help the layout of our presentation slides, as it has better visuals, images, and interactive layouts

MEETING 10

MEETING HELD ON MONDAY, 12TH SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. New canvas slides were created over the weekend
2. Brief presentation was done to explain the slides

MEETING 11

MEETING HELD ON WEDNESDAY, 14TH SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Several improvements were done in the notebooks to help improve the overall scores. These improvements include:
 - Further feature engineering using the variance inflation factor (VIF) method
 - Log transformations of the target variable

MEETING 12

MEETING HELD ON WEDNESDAY, 16TH SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Several suggestions were made to help improve the R2 and RMSE scores:

- Measure the MAE and MAPE performance score metrics
- Calculate the annual child-supporting monetary values by multiplying the monthly child-supporting value by 12.
- Calculate the aggregated child-supporting values by multiplying the monthly child-supporting values by the total number of children each individual supports/ finances
- Hyperparameter tuning of the models
- Use the training data as the X_train and y_train and the testing data as the X_test and y_test. Then, the validation is used to measure the overall performance of the models

MEETING 13

MEETING HELD ON MONDAY, 19TH SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Each team member was assigned to work on a single regressor model and present results by Friday, 23rd September 2022

- Daniel: Super Learner Model
- Stella: Random Forest Regressor
- Obinna : CatBoost
- Ubasinachi: AdaBoost
- George Kibe : XGBoost
- Jeff Ouma: Deep Learning

MEETING 14

MEETING HELD ON WEDNESDAY, 21ST SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

- To finish up on the notebook and push the final notebook(s) to Github
- To add the EDA visualizations to the canvas presentation slides
- To begin researching how to write the white paper report
- To merge the individual investment variables into one single column - in a bid to improve the RMSE and R2 scores

MEETING 15

MEETING HELD ON FRIDAY, 23RD SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Notebooks finished
2. RMSE and R2 scores improved
3. Start preparations for presentations to external shareholders
4. Begin preparing the canvas slides to showcase the progress of the team

MEETING 16

MEETING HELD ON MONDAY, 26TH SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Internal mock presentations within the team's meetings
2. Team advised improving the EDA visualizations embedded in the slides
3. Use case-by-case plots to showcase the EDA of the dataset, instead of plotting all histograms for all columns

MEETING 17

MEETING HELD ON WEDNESDAY, 28TH SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Second presentation with external shareholders held
2. The team discovered an oversight in the target variable, i.e. we had been using the wrong target variable all along
3. Discussions to eliminate this error and use the correct target variable
4. Discussions to improve the canvas presentation slides to showcase the team's progress

MEETING 18

MEETING HELD ON FRIDAY, 30TH SEPTEMBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Third presentation with external shareholders held
2. Team used the correct target variable
3. RMSE and R2 scores greatly improved to be below the 3000/ 0.95 targets
4. Team also used the MAE and MAPE metrics, as earlier advised
5. Team advised by Willie to check on the overfitting of the regression models and improve it

MEETING 19

MEETING HELD ON MONDAY, 3RD OCTOBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Meeting was held to discuss the performance of last week's presentation
2. Great improvement noticed in the format of the presentation slides as well as the models' performance metrics
3. Discussions on how to decrease overfitting of the models, through better scaling
4. Agreement that Ubasinachi shall work on the team's white paper
5. Agreement that Obinna and George shall work on the deployment of the model in the cloud, specifically using AWS resources and Streamlight

MEETING 20

MEETING HELD ON WEDNESDAY, 5TH OCTOBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Agreement that one way to reduce models' overfitting was to use scaling only on the training dataset and transform the test/ validation datasets
2. Another technique is to use the LightGBM model, as opposed to the XGBoost model

MEETING 21

MEETING HELD ON FRIDAY, 7TH OCTOBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Nomfundo to send us the white paper template to our emails
2. Everyone to input their share/ contribution into the white paper template
3. To proceed with the models' cloud deployment

MEETING 22

MEETING HELD ON MONDAY, 10TH OCTOBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Data engineers experiencing challenges with successfully creating an AWS Key-Pair, hence model deployment is at a halt
2. The First draft of the white paper created
3. To have Nomfundo go through the first draft of the white paper

MEETING 23

MEETING HELD ON WEDNESDAY, 12TH OCTOBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Data engineers collaborating to find a way to successfully create an AWS Key-Pair
2. Nomfundo went through the first draft of the white paper and requested us to make several edits and additions to it, e.g. add more data visualization plots

MEETING 24

MEETING HELD ON FRIDAY, 14TH OCTOBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Team members had a retrospective meeting to discuss the challenges they experienced throughout the week
2. The internship supervisors, i.e. Gabriel and Nomfundo, informed us that there shall be a meeting on Monday (next week on 17th October) to discuss how to move workloads to AWS and S3 buckets
3. Nomfundo requested the team to create a better design and layout for the white paper report

MEETING 25

MEETING HELD ON MONDAY, 17TH OCTOBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. Data engineers have now succeeded in creating an AWS key-pair feature and have also managed to move several workloads to the cloud, e.g. the notebook,
2. A new design and layout of the white paper report have been created
3. The white paper report has also been edited further to include more data visualization plots as well as better documentation about the model's deployment to the cloud. The white paper is now complete and ready to be sent to the appropriate entities for marking

MEETING 26

MEETING HELD ON WEDNESDAY, 19TH OCTOBER 2022, AT 10:30 AM SOUTH AFRICAN TIME

Matters discussed:

1. To finish working on the presentation slides as well as the minute's documentation, then send them to the appropriate entities for marking, by Friday 21st October
2. To start practising on how to do mock presentations and the final presentation on Monday, 24th October

