# Progress report for the data team

1. Identify training needs for the data team
   1. Identify short courses

**Comment: we have been learning as a team from various online sources mostly data-related courses. We developed a learning path for each team member based on their competence and have tackled all the various key learning outcomes for the year.**

* 1. Write a plan on how to learn for each project

**Comment: we have been learning from each project mostly focused on automation (from data processing to consolidation and cleaning) and data processes standardization (to get consistent data outputs across various projects)**

* 1. A plan on what we could experiment with the per project.

**Comment: we have been experimenting with new technologies such as databases/warehouses, google sheets and various dashboarding technologies for optimal work.**

1. Hold a meeting with Anne every other week for half an hour

**Comment: this is happening and we actually now meet weekly instead of the proposed biweekly meetings**

1. Provide specific data brief for each project

Comment: **We have been identifying the tasks required, some indicators created and timelines have been provided as well as client needs but mostly to the data team.  
weakness/solution: I will be preparing these in an email write-up and sharing them with Anne and the team members instead of verbal/telegram expectations.**

1. Work on ECD in South Africa

**Comment: this has started. Faith and I have set up biweekly data consolidation pipelines (done) and have started analysis of the data although this is in the early stages.**

**We had a meeting with Mhlalisi we don’t have good timeline and goals yet**

1. Learn and experiment for 4 hours each week (not related to work)

**Comment: This is ongoing although it has not been possible most weeks due to challenges to do with time and competing projects.**

1. Work on the corner shop project

Comment: **Comment: this has started. Faith and I have set up biweekly data consolidation pipelines (done) and have started analysis of the data although this is in the early stages but have made progress in the setups.**

1. Plan a format on the Infinity board with Mahlet

**Comment: we had the meeting with Mahlet, yet the data team hasn’t started using the infinity system to assign work yet.**

1. Build systems/standard procedures for all projects

**Comment: This is ongoing but has had some successes**.

1. **We created data manually for onboarding new team members and explaining all the processes.**
2. **We have developed standard consolidation (inspired by NYU) for finbit data to ensure standardized data formats**
3. **We also have reports templates for the data as well as a defined data pipeline that ensures data moves seamlessly from finbit to reports/visualizations**
4. **We now have a code management /backup system (GitHub) that we use to manage code versions as well as monitor work progress as well as act as our backup system in case of a lost computer /or lost code.**
5. **We have also started developing guidelines and data documentation for each project to act as a guide**
6. Work on credit scoring based on finbit data

**Comment: This is still ongoing we have the following achievements –**

1. **We have developed a preliminary model on the Stachanov system based on SFD data**
2. **Have had several**
3. **We have developed some interesting variables that we base the credit scoring on**
4. **We have had some case studies where we recommend some firms for credits coring such as plus-plus loans**
5. **We have good clean data from the SFD for the credit scoring that we are currently doing a deep dive into to get a better**
6. Send Anne Marie past reports

**Comment: I will confirm if I sent some past reports for Anne’s comments. I did send some past blogs and Anne already shared her comments. I shared past blogs with Anne and she reviewed them and gave comments.**

1. Share and write blogs

**Comment: I have only shared one instead of the 4 that were supposed to be made. I should work on this to ensure more blogs in the future.**

1. Break down projects to tasks and share these with the team members

**Comment: currently ongoing and progressing well**

1. Choose and develop technologies for data and explicitly continue to use a range of technologies

**Comment: we have been incorporating data technologies into the data process some are as follows.**

1. **Database, we have experimented with using databases for the data team. The idea is to process data to a clean format for each project and add it to a database, then use the clean data for further analysis and reports. This is in sharp contrast with past usage of data where we clean only the parts of the data we need for the report and then stop there.**
2. **Google sheets. This has been working well in place of databases as they are a simpler technology. So far, we have been sending clean data to google sheets and sharing it with the clients or doing further analysis from here. These are however slow for large data and in the future, we shall use them in conjunction with databases.**
3. **Data pipelines. We have been developing modern data pipelines for the lift data. This is where we automate data progress from finbit all the way to the reports. When new data comes its automatically picked, processed and saved as well as analyzed. This will continue to progress as we get new technologies.**
4. **This year we introduced dashboards. We are using google data studio and plotly-dash as our technologies of choice but have been experimenting with tableau as well as PowerBI**
5. **Machine learning and AI. We have implemented some machine learning models on the Stakhanov system and also in python for the credit system. We plan to hold more discussions with the IT team for other models they may need in the future and start working on these as well.**
6. **Others. We also have new minor technologies we have been implementing, such as new python tools for analysis, R-tools such as shiny, data applications on Heroku etc. some are in the experimental stages, however.**
7. **Data backups. We have developed a data pipeline for backups mostly GitHub as well as google drive for the data team and have shared the logins with the technology manager (Haimanot).**
8. Prepare infographics for each project starting with the ECD project.

**Comment: This didn’t happen for any project this year. I plan to implement this in conjunction with the blogs. We have started extracting data from past projects and cleaning it with the aim to generate new insights, creating blogs based on these insights and then incorporating infographics and dashboards for these. Each team member has been assigned an old project to start looking into and preparing as follows:**

1. **john - 3W /NYU project (started)**
2. **Nobert – rise data (started)**
3. **Cosmas – Hekima data (to start January)**
4. **Sheilla – (Yar project) (started)**
5. **Lynn – SEEP (to start January)**

**The outputs should be clean data, some infographics, and a blog**

1. Create tasks in infinity for all items that we have in the weekly call monitor spreadsheet

**Comment: So far haven’t been adding our weekly talks to infinity but we have been monitoring the progress on google docs.**