# First report of Open Banking Data Analysis Project

Team 45 11/24/2019

## Updates from the first meeting with clients:

We will be working with NTTDATA to develop a webapp that focuses on the idea of open banking and gives consumers forward looking insights on their spending. Clients also specified technologies that were to be used in this webapp. They include MarkLogic, a NoSQL data hub; HyperCube, a data modelling tool; and python or C++ feather for the frame of the webapp. And here are the detailed notes and requirements from clients:

Notes on Open banking:

- Open Banking tries to open up banking data to consumers to allow for faster establishment and more useful data to be extracted
- Open Banking says all these banks should have an API so that a third party (could be anyone) can access ad extract all the data, but consent has to be given so that personal data can be given
- However, we will not need to really get access to open banking API, we will only be storing sample data to our database.

Notes on data:

- MarkLogic NoSQL database will be used
- We need to get Azure set up and put Developer version of MarkLogic into Azure and give access to clients
- Clients required us to use Hypercube for data modelling

Output requirements:

- Data will be used for Business intelligence
- Step 1: Give consumers basic overview includes visualization of categorical spending
- Step 2: Use R to provide forward looking insights ( build something like a basic flag, e.g. is the person gonna struggle financially.) for personal financial management.
- Use Python framework to build up a web app and get a dashboard from themeforest or smth

## Progress so far:

We have set up Azure database and gave our clients access to the database.

For DevOps:

- We have also used Azure for DevOps in our development. Using the webapp services on Azure, we have created a github repository for our WebApp. As this repository is linked to Azure, the WebApp will be updated and deployed automatically and this allows our development to be more efficient.
- We hope to know how clients carry out DevOps and try to cater to their working style

We have also split tasks into three parts and each teammate has been assigned different tasks during the first sprint.

## Virtual machine and SSH set up (taken care by Raghib)

Progress so far:

- We created a Virtual Machine and deployed a MarkLogic database on it, and then we figured out how to “ssh” into this virtual machine.
  Note: The first virtual machine we created had to be deleted since it was set up on the wrong subscription, and we were running into problems so we were then given a subscription on which we deployed a new database.
- Also found a software called “WinSCP” for windows which allows visualisation of the virtual machine- you can drag and drop files between your computer and the virtual machine, this will make working with data easier.
- Also created the website we need to make as part of the deliverables in Jan.

Next steps:

- Will continue working on (deliverable) website, updated with requirements, project abstract and DevOps.
- Need to start the front end of the open banking web app we are developing, figure out how we are going to link this to a python framework (which Lib will work on).

## MarkLogic database (taken care by Yuheng)

Progress so far:

- We have joined MarkLogic university and started to follow the courses to learn about how NoSQL and especially how MarkLogic works
- We have installed MarkLogic Servers and Converters on both Windows and CentOS ( as windows is only for development and not great in servers)
- We have set up the first instance of MarkLogic on localhost

Next steps:

- We will set up the databases and wait for sample data to be put into the database
- We also need to look into Data Hub 5.0 which is used by clients and help build and manage databases
- After databases are set up, we need to look into APIs provided by MarkLogic that allows our middle tier (most likely Python) to talk to the databases.

## Open banking data (taken care by Lib Kai)

Progress so far:

- We've looked into Open Banking API to try and get a hold of sample data but after reading through guidelines and documentations, we realized that we had to be a part of a legal entity and submit legal documentations to sign up and access the API.
- Therefore we're currently looking into alternative open banking APIs such as tink and openbank.

Next steps:

- Find an Open Banking API that we can access sample data from and start importing it into our MarkLogic database.
- Read up on HyperCube data modelling and understand what we're required to do.