# **MEETING MINUTES WEEK 2**

MEETING MINUTES - 2 EY

DATE

AGENDA

1. Brief program summary

2. Problem statement

3. Team structure

4. Business requirements

5. Design principles

6. Ways of working

7. Grouping activity

8. Next steps

EY WEEK 2 - Meeting Minutes - Team 4

DATE: 6/08/2024

**AGENDA**

1. Roles + Discord set-up
2. Initial thoughts + questions on the topic
3. Feasibility report
4. User stories
5. Separating ourselves into two groups
6. Next steps

**KEY ACTIONS**

* Create Options Analysis for MVP and technologies used by each role
  + All team members
  + By next f2f meeting
* High level Solutions Design
  + Main Alan but whole team
  + By next f2f meeting
* Learn PowerBI and create user stories
  + All members
  + Throughout the project
* Create github
  + Alan
  + by end of week
* Change Feasibility Report based on feedback from EY
  + All members
  + By Thursday
* Finalising project management tool
  + Noor
  + By end of today
* Join the discord and google drive
  + EY mentors
* Send presentation pack
  + EY mentors

**KEY OUTCOMES**

* Deliverable 1 finished and submitted by Thursday
* High level Solutions Design and Options Analysis presentation done by next meeting with client
* Github created
* All member familiar with Power BI software

**KEY DISCUSSION**

* Provide alternatives **(options analysis)**

o Different MVP options (if everything goes well, or if we don’t have enough time etc..)

* Complement the other team
* Initial showcase of what we are doing

o BA coordinate everything (different technologies needed for each role); costs, processing per minute/per day (fastest and cheapest)

o **High level solutions design** for the project

o See what we are thinking about

* **NOTE:** Fix the feasibility report the alternatives sections

o what azure, cloud

o data processing

* Focus only on Motor Fraud claims data
* IDE - think ourselves
* Build the github
  + Don’t be afraid of not knowing thing
* We don’t need to create No queries or ad hoc .. for the NMRA BA’s
* DATASET
  + 10,000 lines of data initially
  + Think about finding datasets or creating data but try to find data first
  + Have a good stretch period of data 5 to 10 years
* Discuss with other team if we are going to write examples of what is fraud in the data set fraud or the ai team is going to hard code it
* Inference and data pipeline
* USER STORIES:
  + Mainly more data side
  + Also include the AI side as a client
  + 20 user stories
  + E.g. The agent wants to see this data that can’t be required and the system shows a notification….

HLSD look up

# EY Meeting 3

**Date:** 13/08/24

**Location:** EY, Sydney Office - Room 33.21

**AGENDA**

* Weekly Standup
* Discussion of optional analysis
* Q&A with mentors

**KEY ACTIONS**

* Create pack with AI team for next standup
* Set up a meeting with AI team for discussion and collaboration
* Combine both teams options analysis with costs (Overall cost and solution)
* Gantt chart - combined with AI team (Remember to block out mid sem break)
* Send out scoping document by next week for feedback from mentors
* Prepare questions for deliverable 2

**KEY OUTCOMES**

* Combined Optional analysis including combined costs
* Combined Gantt Chart
* Combined pack to present to mentors

**KEY DISCUSSION**

* **Presented our option analysis** 
  + Recommending option 3 (Databricks option)
  + Received feedback from mentors that we should go ahead with Option 1(Azure function)
* **Important to keep remembering:**
  + How can the AI team integrate their building within our layers.
  + Please regularly review with Ai team before presenting any solutions
* **How our standup structure should be:**
  + What did we do , Future tasks
  + Problems we had this week
  + one person speaks about a team update
  + List of questions we want to go through
* **Stretch goal:**
  + Data catalog as a final product
  + Apply conditions
  + From a pipeline perspective

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# **Meeting 4**

AGENDA

· Sprint 2

· Present combine group 13-14 pack

· Discuss options analysis

KEY ACTIONS

- Set a meeting with another team to practice with the pack and presentation.

- Remove the visualization in the options analysis and send to the client

- Starting with the option cheapest/smallest and then go higher/larger for option analysis

- Better communication

- Start creating user stories

- Combine the both team 13 and 14 Gantt chart

KEY OUTCOMES

- Create combined Gantt chart

- User stories

- Updated options analysis

KEY DISCUSSION

Presentation – be faster and concise within 30 minutes (both sprint and presentation)

Repository:

· So far good

· create bronze and silver folder so you can library stuff out

· repositories get messy so don’t be afraid to have a lot of folders rather than just a small amount of folders

Data Synthesization - 3 techniques mentioned were

· make\_classification function in scikit learn

· GAN (Generative Adversarial Network)

· OpenAI Codex Model. Mentors were fine with them.

Look into:

Chain of thought

Prompt engineering

- Try and be more involved – Ninuri and tash

EY meeting minutes

· Finalizing deliverable 2 draft

· Looking up on data cleaning techniques and methods

· Looking up on coding convention structure for our GitHub

· Editing deliverable 2 to reflect our feedback

· Writing the code convention

· Writing read.me and documentation files in the GitHub

· Helping Adam clean the data

# **EY Meeting 5**

Date: 27/08/24

Location: EY, Sydney Office - Room 34.19

AGENDA

· Weekly Stand-Up

· Combined Presentation with Team 13

· Roadblocks

· Mentor Discussion

KEY ACTIONS

· Put in Azure Application Insights into the Options Analysis by TODAY (Alan)

· Give edit access to the Jira to Josh – today (Noor)

· Make the Gannt chart label the task name – by next meeting (Noor)

· Send live link of document to Josh – today (Tash)

· Get feedback for Deliverable 2 from EY mentors (EY mentors)

· Add Noah’s GitHub to the repository

· Start cleaning dataset

· Research on data synthesisation methods

· Work on next week presentation with group 13

· Look for more data sources

KEY OUTCOMES

· Deliverable 2 edited and submitted

· Final Option Analysis sent to EY

· Revised User Stories Spreadsheet sent to EY

· Josh has access to Jira and the Deliverable 2 document

· Document of MVP/prototype discussion to EY

· Gantt chart edited to reflect feedback

· Noah added to the GitHub

· Code Convention completed

· Next week

KEY DISCUSSION

User Stories:

· The more the better (minimum 50)

· Have in a spreadsheet format

· Change to technical user stories and business User Stories

· Add more business user stories

MVP

· MVP is the dataset

· Anything that we feel that reflects the work we did.

· Discussed what our MVP will look like with the mentors and send a document on our idea

· Don’t normalize data for the first MVP

Deliverable 3

· MVP Prototype 24th September Presentation (maybe)

· See what we think what will be in the MVP and send document and get feedback

Presentation

· Updates on the roadblocks – slides on how we overcame previous roadblocks

· Documentation on the GitHub

Dataset

· Try to find another dataset and merge it.

· Look at our data’s shape for “age”

· Look at Drivers Licence Age statistics (prolly from serviceNSW) then refine (add weighting) to our randomisation for calculating “DriverExperience”

· Potentially put indexing into our database once it’s merged and large.

· Don’t segment/separate data yet as our database is still small.

· Our MVP is basically just the diagram on slides.

· Look at potential errors in our source data and try keep these errors in our synthesize data

· Diversify data sources

What mvp 10000 clean and synthesized data

And send mentors documentation what it will be mvp suggest/option analysis

Don’t normalize data for the first mvp

Diversify data sources

# EY MEETING 6

Date: 3/09/24

Location: EY, Sydney Office - Room 35.14

AGENDA:

* Weekly stand-up
* Mentor discussion

KEY ACTIONS

* Create discord chat for uploading documents for mentor feedback (Alan)
* Complete imputations methods (Tash and Ninuri)
* Send user stories to Josh (Tash)
* Fix dataset according to mentor feedback - perfect/clean dataset (Adam and Noor)
* Start working on Deliverable 3 (whole team)
* Work on next week weekly stand-up

KEY OUTCOMES:

* Updated dataset structure
* Draft of imputation methods

KEY DISCUSSION:

* Explained options analysis to Josh for both teams and MVP proposal has an initial approval
* Discussed structure of Github repository and workflow with Noah
* Discussed data pipeline structure with Noah
* Discussed data lake structure

Deliverable

* Try and submit earlier
* Leave more time for formatting and editing
* Don’t be afraid to follow up on feedback

Dataset:

* Discussed about the structure of the datasets
* Discussed about perfect datasets/clean datasets
* Discussed about biases in data
* Enriching vs synthesizing

# EY MEETING 7

Date: 10/09/24

Location: EY, Sydney Office - Room 33.09

KEY AGENDA

* Weekly stand-up
* Mentor discussion

KEY ACTIONS

* Work on fixing the accuracy issues within random forest classifier (Tash)
* Work on cleaning dataset 3 (Whole team)
* Work on increasing the number of rows of the dataset with synthesisation (Noor and Adam )
* Work with the AI team to help them ingest the data (Alan)
* Search for sythesisation method 2 (whole team)
* Work on deliverable 3 (whole team)

KEY OUTCOMES

* Send 6.5k rows to the AI team by next week
* Send 20k rows to the AI team by next week
* Imputed missing values in dataset 3 as soon as possible
* Clean dataset 3 by end of this week
* Determine sythesiation method 2 by Friday

KEY DISCUSSION

Random forest:

* Issue: accuracy is low (42%)
* Try to train the column on missing and complete and not just the 80:20 split as currently the test values are small as rows with null values are dropped.

Dataset

* Think about the ratio of what is null and not null
* Decide on what columns need to be complete instead of completing all columns due to bias from ML models
* Don’t focus on complete datasets
* More we impute missing values the more the accuracy of the dataset decreases
* Discussed the fraud column (is it based on the data or generated by the claim agent)
* Discussed about the generated columns
* Focus now on just using 1 method to impute missing values as we need more data for the AI team soon do other methods later

MVP

* Communicate and ask what format AI want the dataset
* Gold stage - is a dataset in the format they want - vector/text base
* MVP is for AI to ingest the data
* Steernow for them to ingest the data and not focus on the imputing missing value
* Mvp 1 defining the structure is good
* The dataset must work with their MVP

Plan

* Tomorrow give 6.5k rows to the AI Team
* By the 17th have 20k rows to the AI Team
* By the 24th have 50k rows to the AI Team

# EY MEETING 8

Date: 17/09/24

Location: EY, Sydney Office - Room 33.09

AGENDA

* Weekly stand-up
* Mentor discussion

KEY ACTIONS

* Create 20,000 rows of data for AI by tomorrow (Noor and Adam)
* Finalise cleaned dataset 3 by today (Tash and Ninuri)
* Work on Deliverable 3 draft and send to EY by Thursday or Friday (whole team)
* Work on MVP presentation (whole team)

KEY OUTCOMES

* 20,000 row sent to AI team
* Dataset 3 cleaned
* Deliverable 3 draft sent to EY
* MVP presentation and speech completed

KEY DISCUSSION

Dataset

* Have a threshold for the accuracy and even if its at the bottom of the threshold it will be good
* Ensure synethesisation accuracy is within the set range.

Presentation

* Demo - folder structure, tasks with the most time and effort , walk through data.
* Make the development pipeline diagram left to right
* Emphasise value to the business
* 10-15 minutes then questions after
* Live demo of what we did from left to right
* The presentation is more for the mentors to ask questions and more for them – to fill in the gaps in knowledge
* More causal not too formal
* Be formal - dress formal
* Treat the mentors like they are from NRMA - like a big update - haven seen for 8 weeks
* Can be technical but also has to explain in non-technical terms - business terms
* Restate the problem in one sentence. Then find away to explain how we got to our sub-problem. Build a dataset that is good enough for the AI to team to train the ML
* No introduce the but include for the big presentation
* What’s the problem and solution and how we did it
* Logical reasoning or value

# EY MEETING 9

Date: 24/09/24

Location: EY, Sydney Office - Room 33.21

KEY AGENDA

* MVP 1 Presentation
* Q&A
* Mentor discussion

KEY ACTIONS

* Look through mentor feedback for MVP 1 presentation
* Make changes to Deliveable 3 with mentor feedback
* Work on Deliverable 3
* Format and edit Deliverable 3
* Set up Azure infrastructure
* Look into AutoML

KEY OUTCOME

* Deliverable 3 is completed and edited

KEY DISCUSSION

* Try looking to using AutoML
* Fix presentation slides
  + Diagrams
  + Font
  + Colour scheme

# EY MEETING 10

Date: 01/10/24

Location: EY, Sydney Office - Room 33.21

AGENDA

· Weekly stand-up

· Mentor discussion

KEY ACTIONS

· Set up Azure infrastructure

· Start working on Deliverable 4

· Look into MVP 2 option

KEY OUTCOME

· MVP 2 proposition completed by next stand-up

· Azure infrastructure is setup

KEY DISCUSSION

MVP IDEAS

· ER diagram

· User interface

· Assesses approaches and why we didn’t take them (auto ML)

· Enrich the pipeline

· See if we need more data for the AI team

· What’s the requirements for an LLM model

FINAL PRESENTATION

· Create another table that will have current claims and it will automatically take the claim and paste it to the AI’s input

· Click a button and fills in the chat input

· Create a web application

INSURANCE CLAIMS

· Think about whether the claim will be going directly to the database or whether the claims officer will be inputting the data to the chatbot and then it will go into the database as a feedback loop

· In real life there will various tables such as historical claims or current claims and when someone with a previous fraudulent claim enters another claim it is flagged, they did a fraudulent claim before

**FEEDBACK**

**Presentation feedback**

· Liked how we broke doen the process

· How we did it and why

· Collorative process with the metors

· Explained reasoning very good

· Problems we faces and the steps we took

· Looked at other options considered

· Consider auto ML and why it my not be good

· Consider othe options for final presentation

· Individually good presentors

· Good live demonstration

Improve

· Ordering of the slide

· And the speaking

· A bit overwhelming and too much information at once

· Storytelling

Prototype

· Gone about a very good

· Way we sourced data is good

· Way we synthesised is good

· Didn’t focus on one dataset found a variety

· Liked pipeline

· And how we addressed data quality

· Github

· Build it onto that azure

· All good no glaring issues

# EY MEETING 11

Date: 08/10/24

Location: EY, Sydney Office - Room 33.19

AGENDA

· Weekly stand-up

· Mentor discussion

KEY ACTIONS

* Send Deliverable 4 draft
* Finish Azure functions setup
* Work on MVP 2

KEY OUTCOME

* Deliverable 4 draft sent to EY
* Azure functions deployed

KEY DISCUSSION

Handover Requirements

* User manual, code
* Think about end-users

Final presentation

* Be ready for impromptu questions mid presentation
* Always answer in terms of business value

User Interface

* Discussed about how to make an user interface
* Recommended Azure Static Web App, Next js, Prisma for the UI

# EY MEETING 12

Date: 15/10/24

Location: EY, Sydney Office - Room 33.21

AGENDA

· Weekly stand-up

· Mentor discussion

KEY ACTIONS

* Send deliverable 4
* Look into developing the UI design
* Add authorisation to the website
* Make SQL database for the AI team
* Put scripts on Azure

KEY OUTCOMES

* Deliverable 4 completed
* Website design upgraded
* SQL database sentto AI team by Friday
* Scripts are on Azure

KEY DISCUSSIONS

Website

* Two actions to take with make it look pretty or/and add authorisation
* Look into Tailwind and Materio ui
* Talk to AI team on integrating the AI model to the website