# TITLE SLIDE

Morena and Tena koutou.

My name is Ken MacIver and I’m representing Group 11. Unfortunately the others couldn’t make it today for a variety of reasons, so it’s just me presenting.

## PHOTOS

This is us: we have Jundi, who is based in China, Tom, who is sitting a test for another paper right now, Mei and me.

## GROUP MEMBERS (DETAILS)

This slide shows our names, email addresses and Open Researcher and Contributor IDs.

## DIAMONDS DATASET

We have chosen the ‘diamonds’ dataset.

It has just under 54k rows, and each row contains ten pieces of information about that particular diamond.

Being so large, with over half a million data points, we’ve found it very slow to process and render.

There are ten variables, nine of which give information about the quality or shape of the diamond, while the tenth gives the price

READ FROM SLIDE

READ FROM SLIDE

## Variables

This is the list of variables, with the red font indicating categorical and black font numerical. x, y, and z are various measure of size. depth is a variation of depth, with

depth – z/mean(x,y) = 2 \* z/(x + y)

## Response Variable

Here are the first fifteen rows of the dataset printed out, which gives you a bit of an idea of it.

## Visualisation – the dataset

From the pairs plot we can see the density plots for each of the numerical variables. Notice that while some resemble a Normal distribution (depth, for example) most of them do not. The Cullen-Frey plot of ‘?????’ showed that this variable has a beta distribution

## Visualisation – pairs plot

## Other things of interest

## Next steps

# Misc for question time

* x = 8 zero values, y = y zero values, z = 20 zero values
* replace first 15 rows slide with corrplot.

Diagram, schematic

Description automatically generated



