# Assessment 3

# Mortgage Approval

Group Number: 01

Group members present (Name, ID):

* Jasleen Jasleen (0734327)
* Arvind Sharma (0730475)
* Harpreet Kaur (0733894)

In data science there are numerous numbers of products that are launched every month by different companies. So, the main question that arise is, “Which product is considered as a good product or a good service by company?” the answer can be that fulfil needs of customer or which is very useful for every individual. But there is also another factor that can be considered as important factor is that data used ethically and responsibly, which means trust of user. In today’s world most of the user do not believe on the data products that they are used because of some harmful ads. So, to build a data product we should follow the rules which helps us to maintain the user trust which is very necessary for future.

There are five framing guidelines that help us in building data products and those are five C’s: Consent, Clarity, Consistency, Control and Consequences.

**Consent:** Consent is the first step in which there is an agreement between the people who are providing their data and the people who are using that data. If there is no agreement between them then there must be trust issues between both but, if they have agreement. It is all about what type of data is being collected and how we can use that data which we collected. As in our project that is Mortgage Approval customer know that bank need their data to check for all conditions that are necessary to fulfil. And everyone knows that its Bank’s policy that they never share their customer data with anyone. For example, when you are filling a survey first, they ask that you want to fill it with ‘yes’ or ‘no’ which means consent is there and they asked you that you want to share your information with them or not.

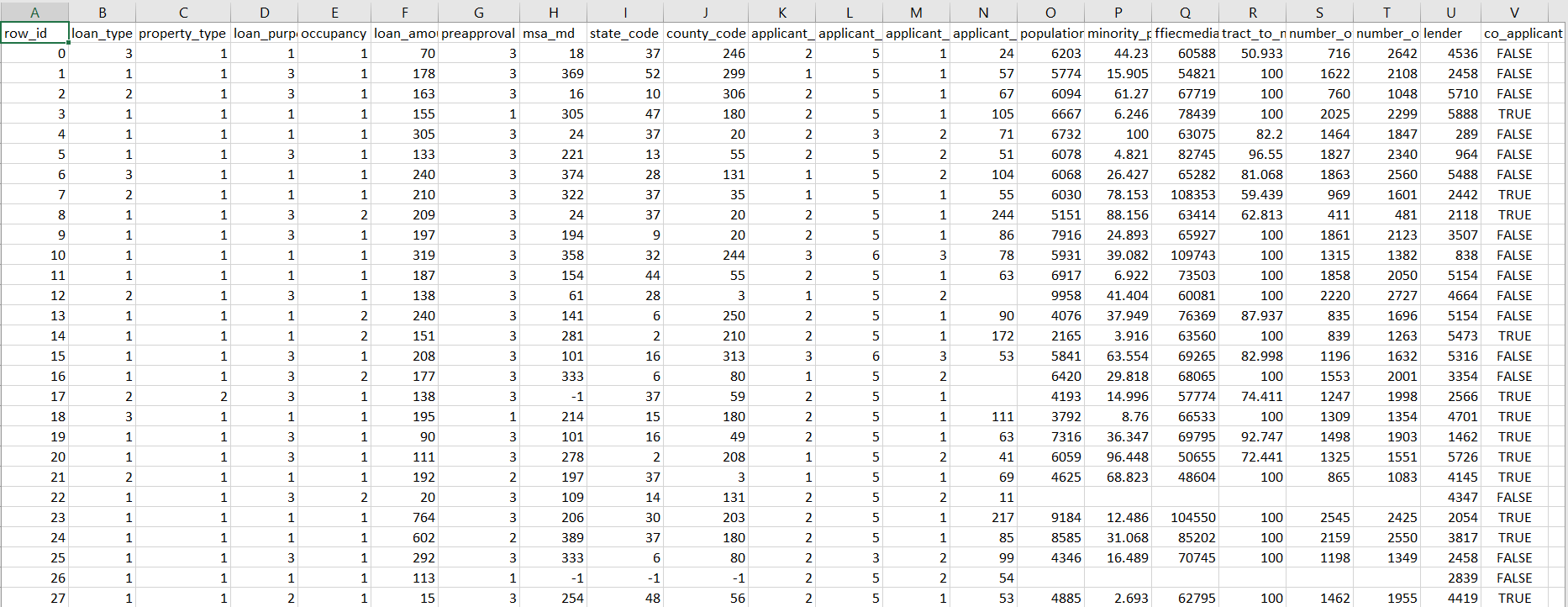
**Clarity:** Clarity is also important factor because is also related to consent. First you must tell clearly everything about the thing you are consenting to. Because user need clarity which kind of data they are providing you and what will happen with that data like how you are going to use that data. If clarity is there, then everything will be good with the product because if we are clear with our thoughts and planning then it is easy for everyone to understand that. In Mortgage approval it is clear that, bank need the data for checking about conditions and bank clearly tell this thing to customers. For example, in any kind of survey, in the staring they mention all the information about that survey so that the public who is going to fill it should be clear about each question.

**Consistency:** Trust is such an essential thing which should not be broken between customer and provider because once its broken then it takes too long to build again. There are many cases in past where customers data have been stolen like: Yahoo!, government data etc. From our project view trust is important from both sides which means if customer need trust from bank then bank also need same thing from customer because in Canada 80% people considered as Bankruptcy which is not good for bank. So, bank also need security from customer that they will not be in loss if any customer will cheat them.

**Control:** It means that the user has a right or not to handle the data which is given to provider that how their data has been used by service provider. But it is totally depending on the service provider because some of them will give them the access and others not. But bank will not provide any kind of access to user, but they give surety that they will not misuse their data and, they have high security control over the customer data which help them to protect the data from leakage.

**Consequences:** Consequences are the results which will come out from the data which may be positive or negative. Because sometimes the data will have some harmful effects which might be not good for customer and other party. So, look at consequences before any side effect is needed. According to Mortgage approval bank should get correct data about customer because it will result in big loss to bank if the data is wrong about the customer. So, they will consider each and every security factor before approving for the loan so that they will be on safe side if there is something happen in future.

**Overlook of the Dataset:**



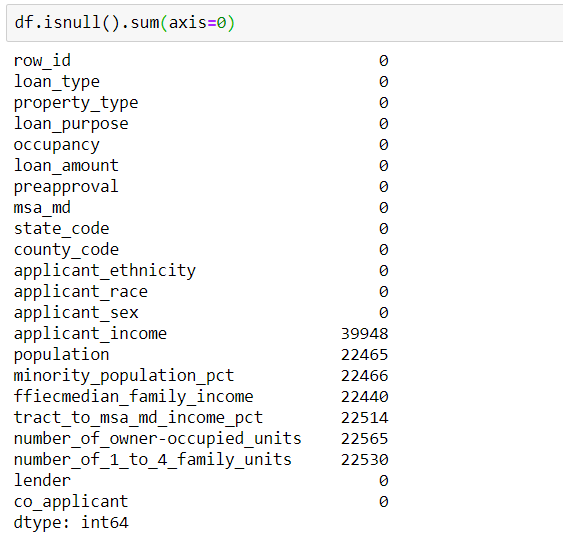
**Why we choose this dataset:** We choose this data because this data has 500,000 rows and 23 columns which is good number of counts to work with and also this data has almost all the information which we want to work in our project.

**Why not other dataset:** Because other dataset might not have this much number of rows and columns and not have this much important information as this dataset must do mortgage loan approval project.

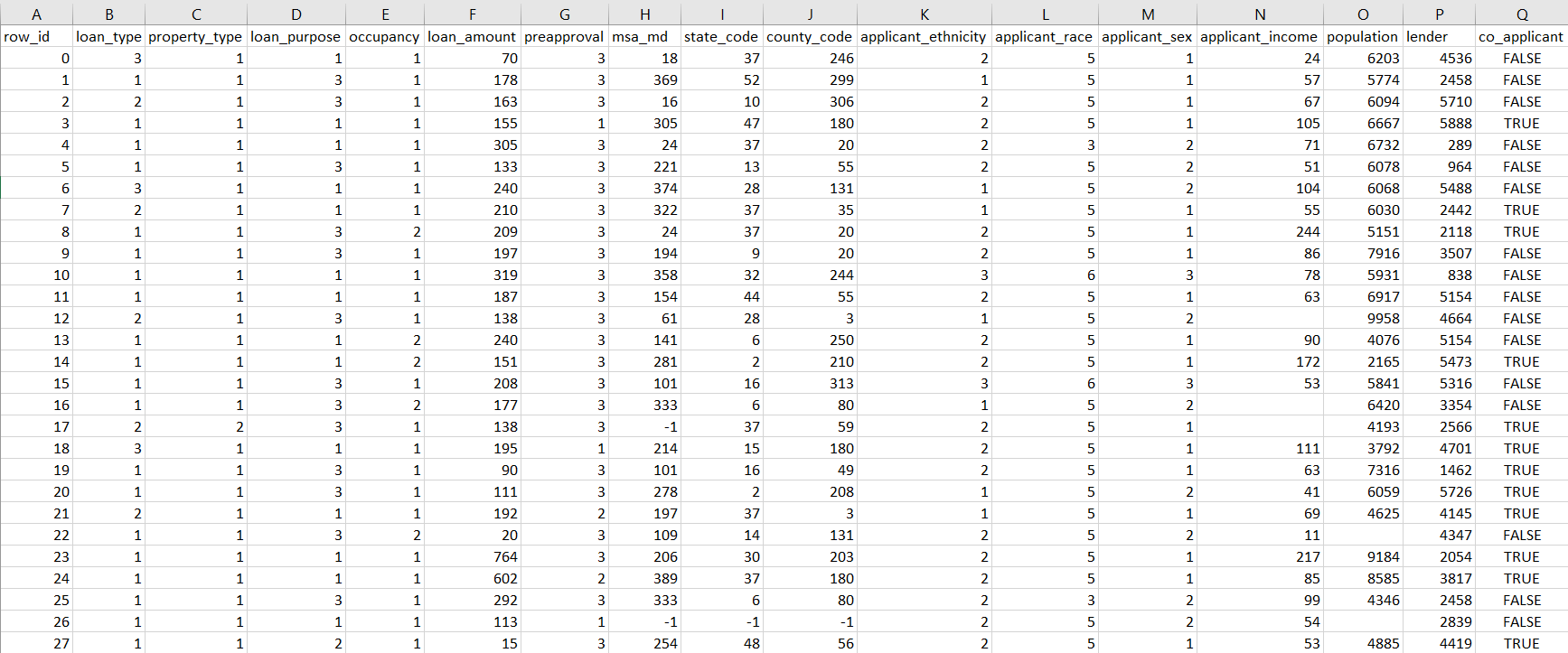
**Different Resources:** We went through various resources for collecting data for our project like we met Scotia Bank associate and we go through various websites for data, but due to privacy and security concern they are not allowed to share the data of their customers.

**Cleaning the dataset:** As it is mentioned above that our dataset has 500,000 rows and there are 174,888 null values in various columns. From them we are going to delete some columns and some of them we will filling by doing some calculation on that.

**Count of Null Values:**



**Removed unwanted columns:**



**Final Solution:** After going through various approaches for getting real data, it is difficult to get that as now-a-day the privacy for each and every company is one of the most important part for their customer. So, finally we get data online from GitHub which is an open data and has 500,000 rows and 23 columns which is enough to perform different models.