Rahul Thaker

**Predictive analysis**

**ANZ**

**Objective:**  Identify the annual salary of the customers based on transactions. And create models to predict the salary.

**Identifying data:**

The Data contains the transaction details of all the customers for 3 months.

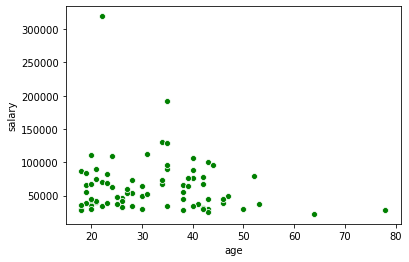
The annual salary for each customer can be identified by using the **‘txn\_description’** column. **Verifying the payment transactions** and calculating the annual salary.

**Given features:**

Age, gender, amount, balance, location for each customer such features are given. Which are not very helpful or related to the salary of the individuals.

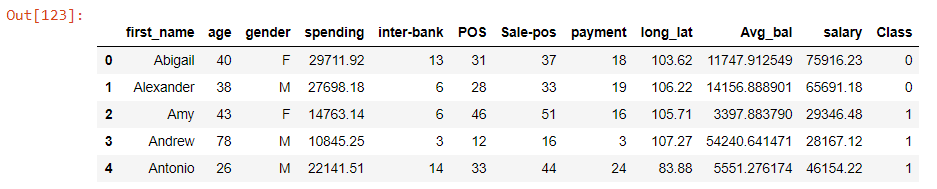
For example:

The relation between age and salary can be explained by the given scatter plot (No linear relation)



**Feature engineering:**

Hence using statistics some features could be extracted from the data. The following table represents all the features with the customers (first 5 records)



The **spending feature** is calculated by using the total amount spent by the customer

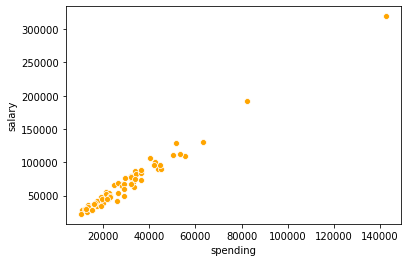
The **txn\_description column is broken down** into its categories and for each customer the total transaction in that category is calculated

**Avg\_bal** column is the average balance of the customer throughout the dataset

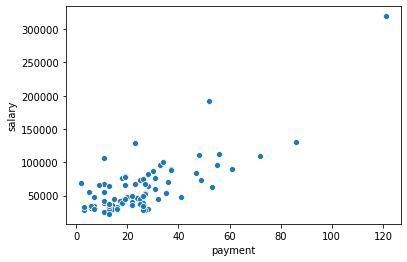
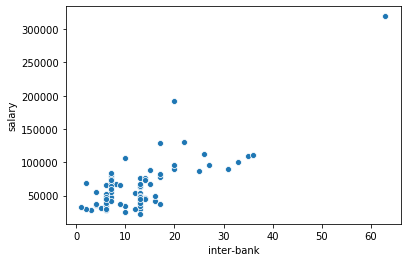
The long\_lat column consists of a number made using longitude and latitude representing a unique number for each location

**Co-relations:**

The spending column has a good relationship with salary as follows: (linear relationship)



The inter-bank, payment transactions also have a very good relationship with salary as follows:



Pos, Spos can also be used while predicting for extra accuracy

**Models:**

**1.linear regression model:**

A multi variable linear regression model could be used to predict the annual salary using the above features

The accuracy of the model : 0.9896187622884146

This model can successfully predict the annual salary of the customers based on the above features.

**2. Decision tree classifier:**

A class column was created which puts a customer in class A if the salary is higher than average salary of all the customers and

Class B if the salary is lower than the average of all the customers.

The decision tree model can classify new customers based on their salary and segment them (new classes can be added)

DecisionTrees's Accuracy: 0.975