Final Presentation

X and Y

- We take age, balance and day as our X and duration as our Y

```
[27] X = data[["age","balance","day"]].copy()

[28] Y = data["duration"]
```

Train Test split and model

- According to our findings from tryouts of support machines regressor and random forests regressor we decided to build a linear regression model which has the lowest error rate. Here is our model and its coefficients.

```
[29] from sklearn.linear_model import LinearRegression

[30] model = LinearRegression()

[31] from sklearn.model_selection import train_test_split

[33] X_train, X_test, y_train, y_test = train_test_split(X,Y,test_size=0.3,random_state=1)

[34] model.fit(X_train,y_train)

     LinearRegression()

[35] model.predict(X_test)

     array([261.63366483, 264.03917216, 267.71435114, ..., 274.24313072,
            258.78076605, 259.82857804])

[36] model.coef_

     array([ 0.11561261, -0.00191192, -0.89782235])
```

Model Performance

- With square root of mean squared error of predictions and our test sample of label, linear regression performs best.

```
[37] predictions = model.predict(X_test)

[38] from sklearn.metrics import mean_squared_error
     import numpy as np

[39] np.sqrt(mean_squared_error(y_test,predictions))

     267.19629183361286
```