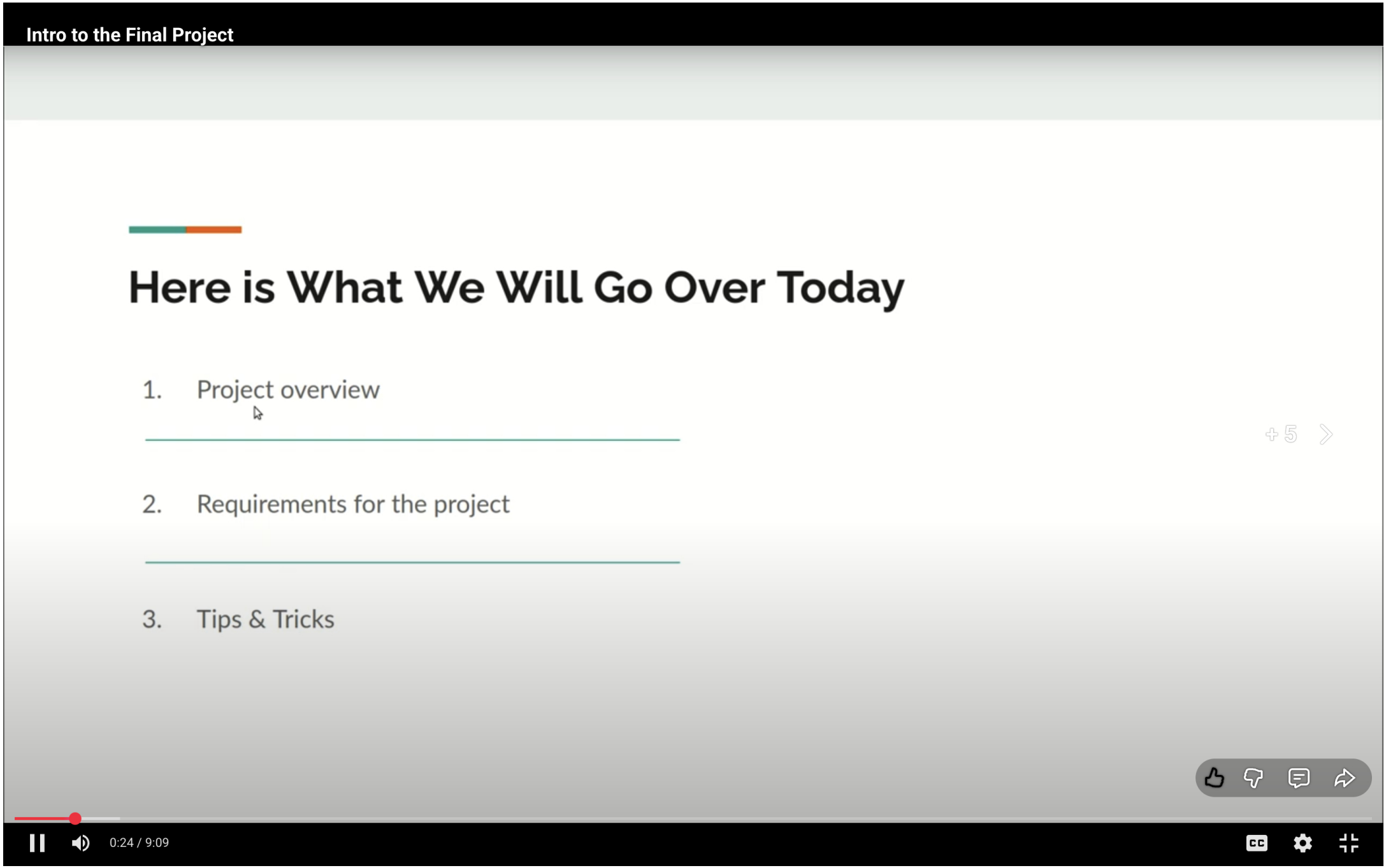
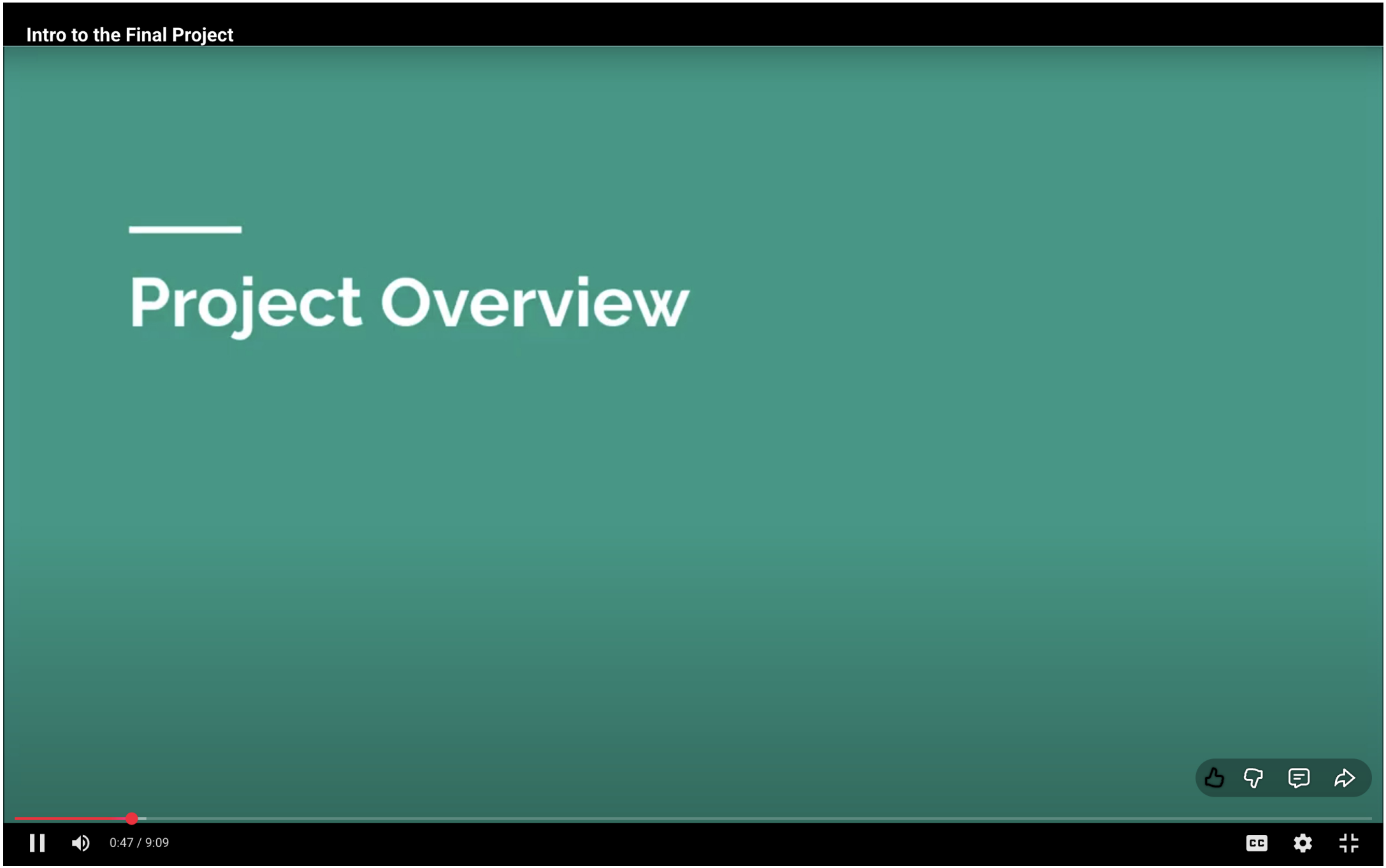
Intro to the Final Project
A Few Words About The Agenda
+ 5
0:16 / 9:09

Intro to the Final Project



Intro to the Final Project



Intro to the Final Project

What are We Going to Solve & Why?

Intro: The telecom operator Interconnect would like to forecast churn of their clients

Business Problem Statement: The company wants to forecast which users are planning to leave

Business Value: To ensure loyalty, those who are going to leave, will be offered with promotional codes and special plan options

+5

0:54 / 9:09

What is Churn?

Churn (or churn rate) is number of customers disconnecting their service over a given time period

In terms of a machine learning problem, our goal is to **classify each client into one of two groups:**

1. Loyal customers who won't churn
2. Customers who will churn in the future

Intro to the Final Project

About the Dataset

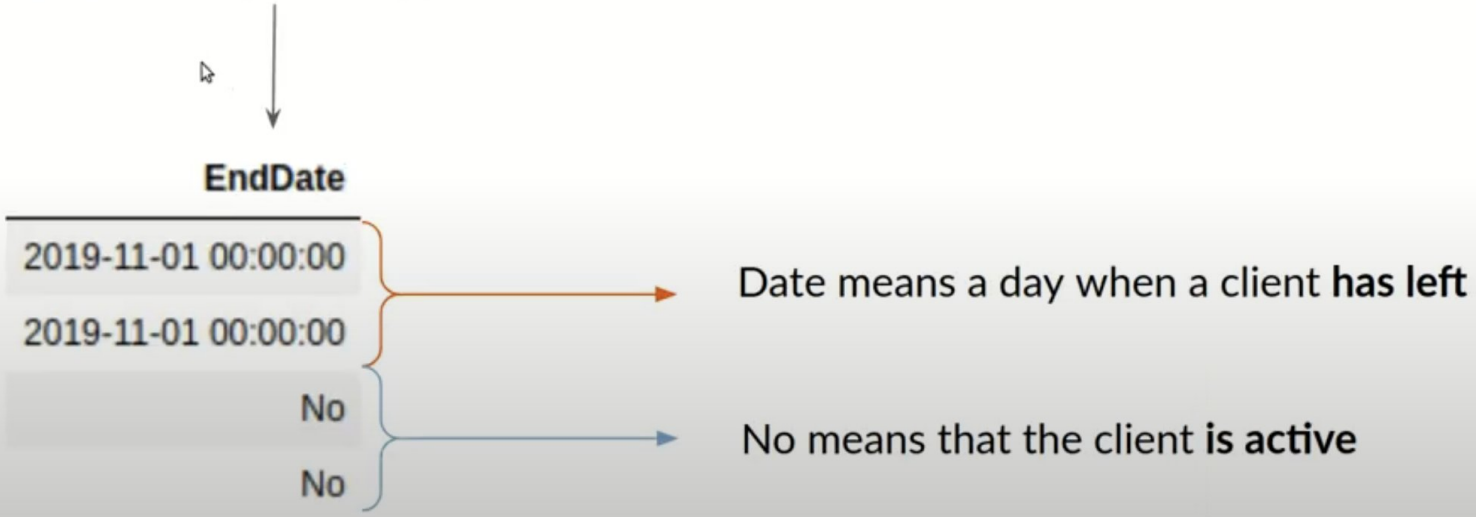
The data consists of files obtained from different sources:

- contract.csv — contract information
- personal.csv — the client's personal data
- internet.csv — information about Internet services
- phone.csv — information about telephone services

In each file, **the customerID column** contains a unique code assigned to each client.

About the Target

Our target is in the **EndDate** column.



Intro to the Final Project

Intro to the Final Project

The Project Consists of 3 Main Stages

1. The project plan

2. The core part

3. The project report

+5

3:25 / 9:09

The Project Plan

A good example of a plan

```
contract['BeginDate'].value_counts()
```

```
2014-02-01    366
2019-10-01    237
2019-11-01    237
2019-09-01    237
2020-01-01    233
             ...
2020-02-01     11
2014-01-01      7
2013-10-01      3
2013-12-01      3
2013-11-01      2
Name: BeginDate, Length: 77, dtype: int64
```

```
len(contract[contract['EndDate'] == 'No']) / len(contract['EndDate'])
```

```
0.7346301292063041
```

Notes about the contract data base:

In the contract we have all the customers id - 7043 total. The data is unballanced (73% not left the program) No missing values.

Task in the preprocessing:

- Need to correct columns name to underscore lowercase.
- change BeginDate date to date type
- TotalCharges change to float64

Task in the features engineering:

- PaymentMethod, Type, PaperlessBilling with one hot encoding
- Create column based on the EndDate says if customer left or not the service.
- Check seasonality for BeginDate (month, day of the week, hour a day)
- After splitting perform class ballancing

Proposed work plan

In the telecom project our goal is to develop a model that can tell is user will leave the service. The steps to achieve that will be:

1. Download the data
2. Explore the data to determine how to treat the data in the preprocessing
3. Perform preprocessing for the data that will include
   - merge the data from all databases to one main dataframe
   - Change column name to consist format
   - convert to desired type
4. Perform EDA to explore the data in depth
5. Feature engineering

A bad example of a plan

Project Plan

1. Download the data
2. Explore the data
3. convert the columns into the right types
4. lowercasing the columns names
5. change the Yes values in the Data frames into 1, and the No into 0
6. Merge the different data frames on the users ID
7. fill in nas that occurred due to the merge with unknown
8. perform EDA (Exploratory Data Analysis)
9. One hot encode the DataFrame
10. scale the DataFrame
11. Create base model for classification: Logistic Regression.
12. test more classification models
13. tune the models in order to get the desired AUC-ROC and Accuracy scores - The metrics that I was asked to measure the models with with.
14. Choose the write model
15. produce conclusions

The Core Part

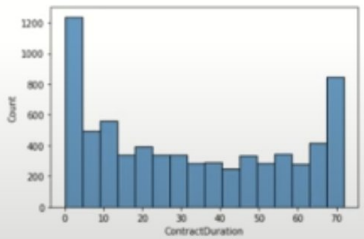
You follow the plan that you have designed.

Your final model should meet the required performance level based on AUC-ROC:

- < 0.75 — 0 SP
- 0.75 ≤ AUC-ROC < 0.81 — 4 SP
- 0.81 ≤ AUC-ROC < 0.85 — 4.5 SP
- 0.85 ≤ AUC-ROC < 0.87 — 5 SP
- 0.87 ≤ AUC-ROC < 0.88 — 5.5 SP
- ≥ 0.88 — 6 SP

The Project Report

1 Your code should be documented



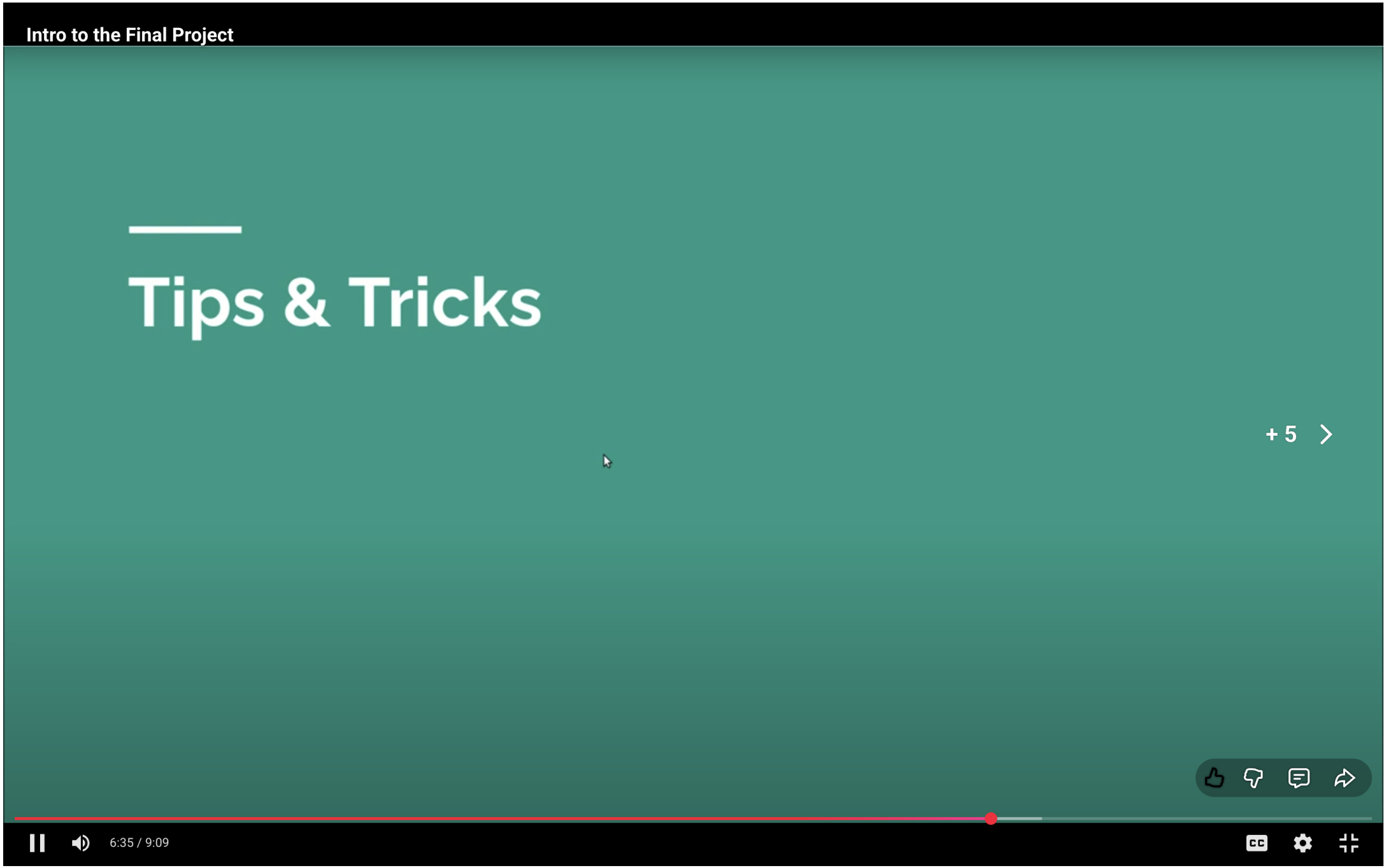
- In the graph above we can see the contract duration for Interconnect operator customers. We can see bivariate distribution- one around 2.5 months and the seconde around 70 months. Ther are two kinds of customers: new ones and some that stay loyal for a long period of time.

2 You need to summarize your work

General Conclusion

In conclusion, I can say that I found several reasons why customers leave the company, as well as possible factors that will help identify such customers. First of all, the monthly payment is such a factor, people who pay less stay in the company longer people who pay more have a higher chance of leaving the company. I also found out that reminding a client of his monthly payment affects whether he stays in the company or not, this is especially noticeable if you consider customers who pay monthly, and the payment method is an electronic check, these customers see the price every month, and also constantly receive paper receipt. On the other hand, people who pay automatically, by credit card or bank transfers stay with the company much longer, I think it's because they don't oversee the monthly payment.

Intro to the Final Project



Things to Consider

1

Feature engineering is a machine learning technique that leverages data to create new variables that aren't in the training set

2

Boosting algorithms seek to improve the prediction power by training a sequence of weak models, each compensating the weaknesses of its predecessors.

3

Critical thinking is the analysis of available facts, evidence, observations, and arguments to form a judgement.