

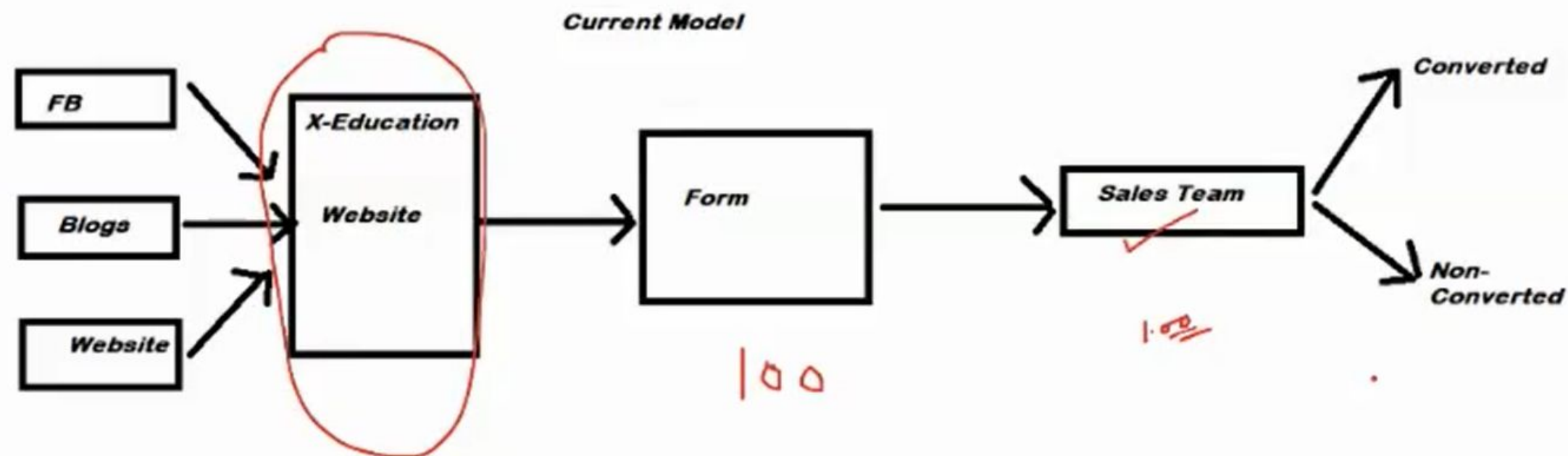
Agenda

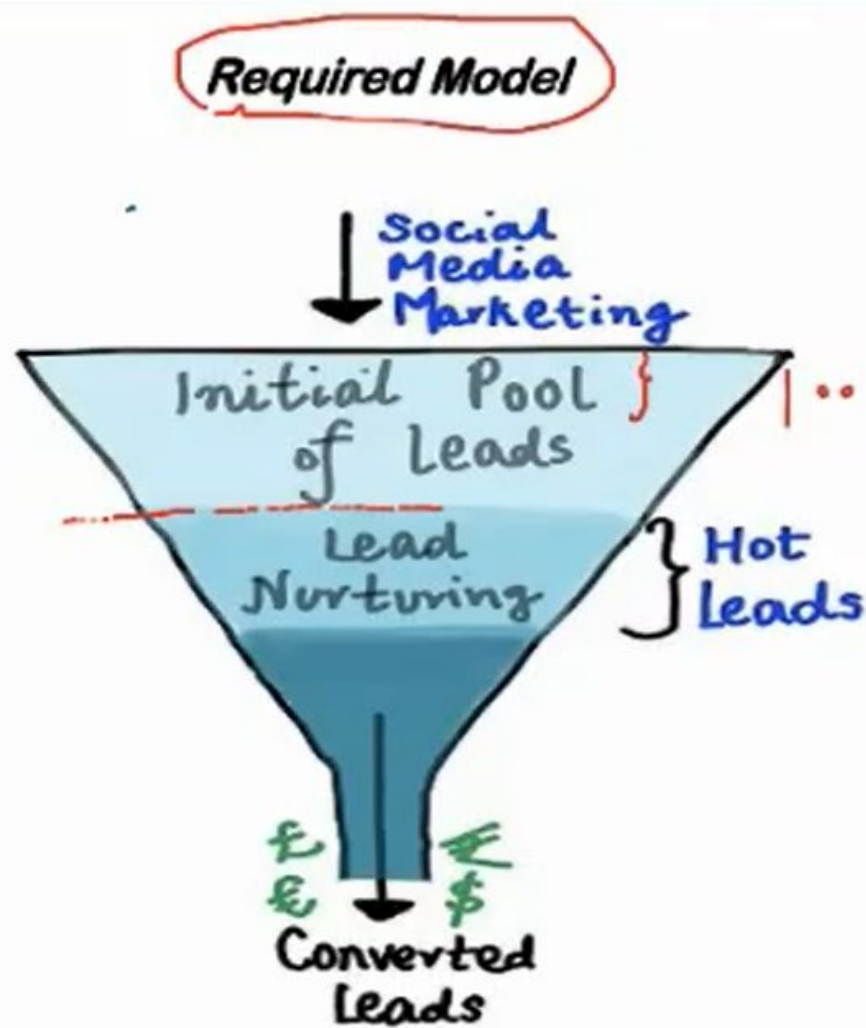
1. Assignment
2. Doubts

An education company named X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.

What you need to do?

- X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.





Data Cleaning:

- ✓ 1. Handle the “Select” level that is present in many of the categorical variables.
2. Drop columns that are having high percentage of missing values. Check all the columns before dropping them.
3. Check the number of unique categories in each categorical column. Here you may need to do something.
4. For the columns with less percentage of missing, use some imputation technique.
5. Finally check the percentage of rows retained in data cleaning process.

<https://stackoverflow.com/questions/20857556/html-form-select-state-and-city>

Made a mistake? [Start Again](#)

Select a city:

Selected State: California

Select a city

Select a city

Los Angeles

San Francisco

Data Preparation:

1. Create dummies for all categorical variables.
2. Perform train-test split.
3. Perform scaling.

{ **Modelling:**

1. Use techniques like RFE to perform variable selection.
2. Build a Logistic Regression model with good sensitivity.
3. Check p-value and VIF.
4. Find the optimal probability cutoff.
5. Check the model performance over the test data.(Confusion Matrix,Sensitivity,F1-score etc.)
6. Generate the score variable.

Submission:

1. Jupyter Notebook: A well-commented Jupyter notebook with at least Logistic Regression model, the conversion predictions and evaluation metrics.
2. Subjective Answers: The word document filled with solutions to all the problems.
3. The overall approach of the analysis in the presentation:
 - Mention the problem statement and the analysis approach briefly.
 - Explain the results in business terms.
 - Include visualizations and summarize the most important results in the presentation.
4. Summary Report : A brief summary report in 500 words explaining how you proceed with the assignment and the learning that you gathered.

ppt (10-12 slides)

- problem Statement
- Assumptions (if any)
- EDA results
- Overall Approach
- Results & Conclusion
- Recommendations