Let's consider the telecom churn example where you have been asked to reduce the high churn rate of the customers. What would be the appropriate data science problem that maps to this business problem and can ultimately help us in solving it?

1. we need to classify the customer segment which are churning out from the non churning customers and identifying what features are causing churning with certain probabilities and choosing those features we try to reduce the churn rate but at the same time consider the tradeoffs for implementation on non churning customers.

Suggested Answer:

We need to build a machine learning model that predicts whether a customer will churn or not. This will enable us to identify features which are crucial in determining churn as well as help us convert those churn customers to non-churn ones.

Approach should be in tune with the hypotheses formulated and should give an idea of your approach of the overall solution i.e. the data that you’ll be using, the EDA that you’ll perform, the ML algorithm that you’ll be using, the Evaluation metrics you’ll be tracking, mapping those evaluation metrics to KPIs and making business decisions.

#### Problem Mapping

Consider the following statement

**When trying to solve a business problem you would always need machine learning models**

The above statement is

False

**✓ Correct**

**Feedback:**

You don't always need a machine learning solution to solve the business problem. A simple EDA analysis may be sufficient in many cases. ML solutions might be difficult to deploy as well as interpret in many organisations. Therefore, always focus on solving business problems rather than utilising machine learning solutions.

**Novelty vs Utility**

Let's say in order to predict the churn rate of the customers you came up with 2 machine learning approaches - Logistic Regression and Neural Networks.

You know that Logistic Regression Models will be highly interpretable and you will be able to identify the important features whereas the Neural Networks model, even though will give a better performance will be less interpretable since it's a Black Box model( it won't explain clearly why it made a certain prediction)

Which modelling technique should you ideally prefer?

We should prefer a practical approach where business problem can be solved with resources at hand and ideally what model make sense for the business should be used. Here we may use logistic regression instead Neural networks where interpretability by the business stakeholders is difficult and we may not have the infrastructure or data to run those neural networks.

**Suggested Answer:**

Recall the business problem here. You need to convey to the client which customers are leaving as well as the features that are more important for their departure. Therefore interpretability matters to them. Hence you should preferably go with the Logistic Regression Model.

The first important step is to build a simple POC or **proof of concept** model. This will get your client excited about the prospects of your solution and also help define the success metrics that you'll be utilising further down the line. It also helps in identifying additional data needs if you have any. As mentioned earlier, there is a significant back and forth going on between the data collection step and the rest of the steps.

brief summary of the EDA steps you're expected to do once you have the data with you:

* **Data staging and clean up**: This is the basic data cleanup and preparation stage. You collect the data from various sources, clean it and prepare the master dataset.
* **Sanity checks**: The next step is doing a quick sanity check of the entire dataset to observe any unusual data points that should not exist.
* **Univariate Analysis**: Finally we begin with the univariate analysis part. This is where visualisation tools like histograms and boxplots come in handy as they help in analysing numerical features.
* **Bivariate Analysis**: Then, you go ahead and evaluate the relationship between the target variable and the rest of the features. Here plots like scatter plots, pair plots, correlation matrices come in very handy to do the analysis. Some segmented analysis can also be done in this step as well.
* **Hypotheses validation**: In this step, you'll be getting some directional insights on whether the hypotheses that you built earlier are showing any promise or not.
* **Feature Engineering**: Finally, if you want, you can do feature engineering to extract useful insights from the given dataset.

Next, we look at the most exciting step of this entire process, i.e. **Model Building**.

There are several other approaches like **CRISP-DM** framework which also help in solving the business problems. You can choose any of the standard approaches to solve the business problem at hand.

Here's a basic overview of the steps that you need to perform:

* Business Understanding
* Develop Hypotheses
* Data Collection
* Problem Mapping
* Solution Approach
* EDA
* Model Building
* Model Evaluation

**Top Three Takeaways from the Business Problem Session:**

1. Understanding a business problem is most crucial step in the problem solving, could be in terms of impact to the business, need to determine what factors help us in solving the business problem at hand.

2. Next important step is about framing the hypotheses and collecting the data to prove or disprove the hypothesis and slowly build a solution and approach which is feasible and practical.

3. Then we do the Data staging, cleanup and sanity check to perform the EDA and then model building and evaluation.

**Industry Case study: X- Ray Machine failure at Hospital – Loss of Revenue and Service to patient**

* **Data Scientist at GE Healthcare**

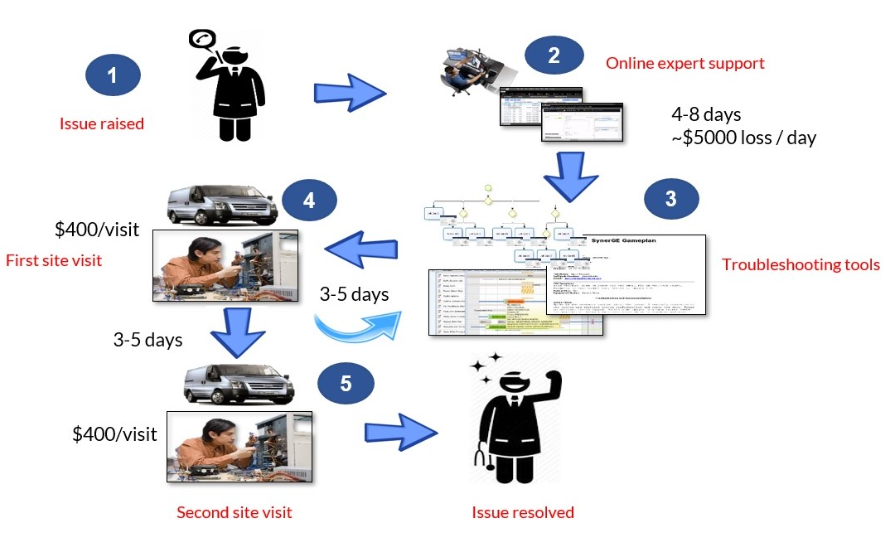
Optimising the Process:

Which steps do you think will be suitable for optimisation using machine learning techniques?

**1. We could predict the failure of the machine before hand by collecting the data and analysing the patterns of the data about the operation of the machine.**

**2. we could reduce the time to bring the x ray machine up again by figuring and replacing the parts which could be lead to failure.**

**3. we could build a model to predict the due maintenence of the machine and reduce the overall cost of repair or damage.**

****

**Suggested Answer:**

We observe that in the third step, the online expert takes around 4-8 days to identify the root cause of the problem and recommend certain possible solutions. Here, the expert looks at the machine logs and the doctor's description and along with their technical knowledge give the necessary recommendations as to which parts might be faulty and need to be replaced. So a machine learning solution would be most suitable for optimising this step where a model can be built on past issue data to give recommendations automatically.