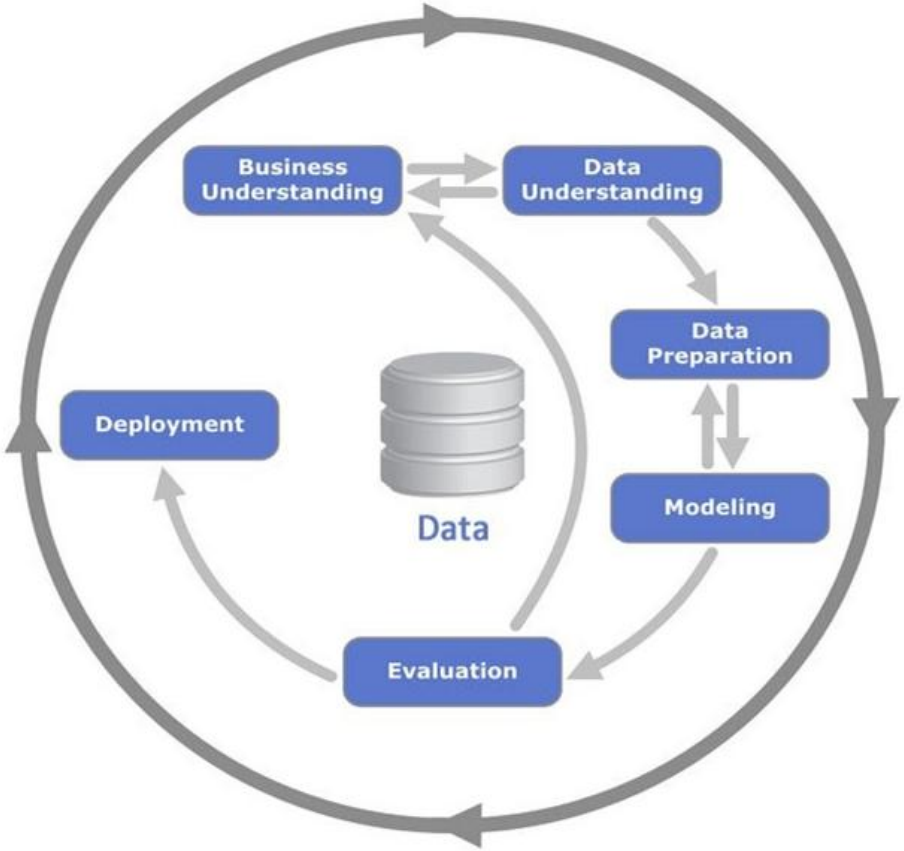
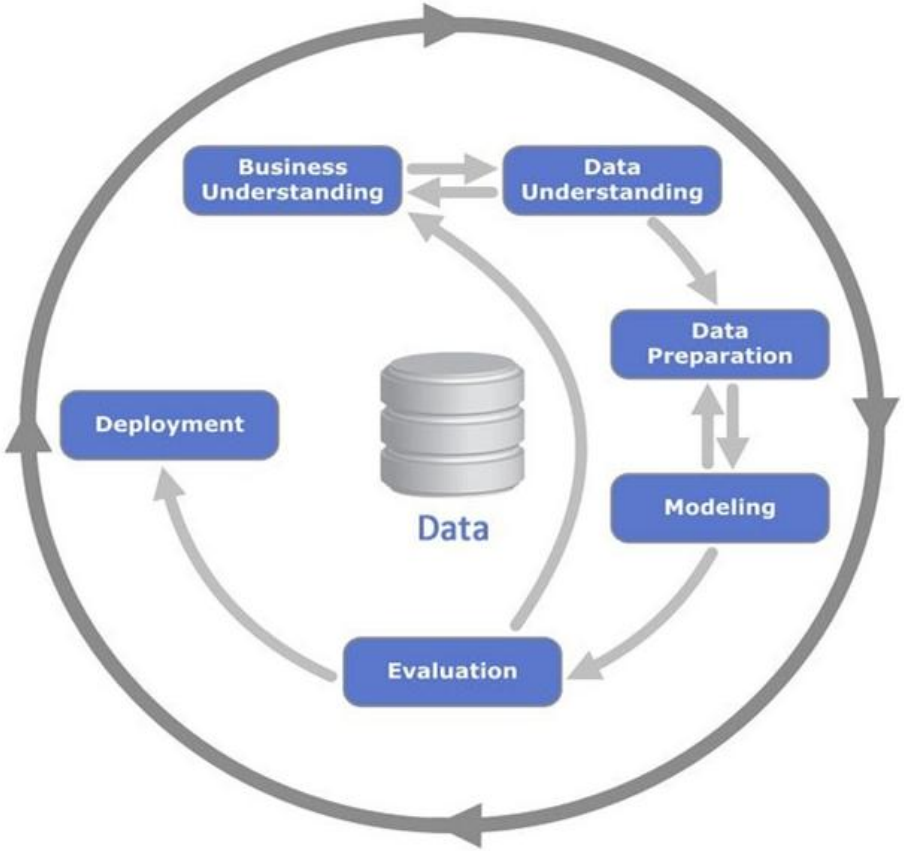
CRISP-DM Framework



Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives *Background* *Business Objectives* *Business Success Criteria*	**Collect Initial Data** *Initial Data Collection Report*	**Select Data** *Rationale for Inclusion/ Exclusion*	**Select Modeling Techniques** *Modeling Technique* *Modeling Assumptions*	**Evaluate Results** *Assessment of Data Mining Results w.r.t. Business Success Criteria* *Approved Models*	**Plan Deployment** *Deployment Plan*
Assess Situation *Inventory of Resources* *Requirements, Assumptions, and Constraints* *Risks and Contingencies* *Terminology* *Costs and Benefits*	**Describe Data** *Data Description Report*	**Clean Data** *Data Cleaning Report*	**Generate Test Design** *Test Design*	**Review Process** *Review of Process*	**Plan Monitoring and Maintenance** *Monitoring and Maintenance Plan*
Determine Data Mining Goals *Data Mining Goals* *Data Mining Success Criteria*	**Explore Data** *Data Exploration Report*	**Construct Data** *Derived Attributes* *Generated Records*	**Build Model** *Parameter Settings* *Models* *Model Descriptions*	**Determine Next Steps** *List of Possible Actions* *Decision*	**Produce Final Report** *Final Report* *Final Presentation*
Produce Project Plan *Project Plan* *Initial Assessment of Tools and Techniques*	**Verify Data Quality** *Data Quality Report*	**Integrate Data** *Merged Data*	**Assess Model** *Model Assessment* *Revised Parameter Settings*		**Review Project** *Experience Documentation*
		Format Data *Reformatted Data*			
		Dataset *Dataset Description*			

https://www.datascience-pm.com/crisp-dm-2/

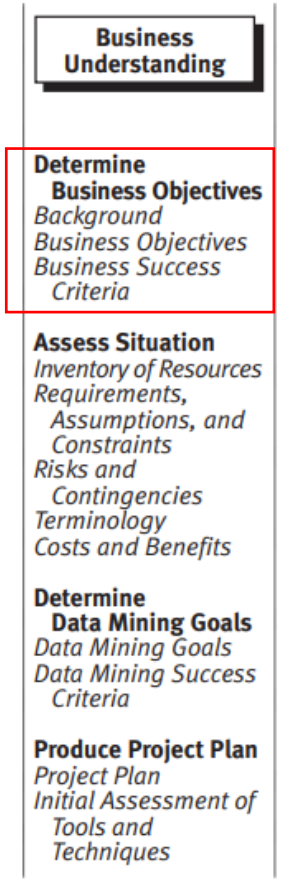
CRISP-DM Framework



Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives *Background* *Business Objectives* *Business Success Criteria*	**Collect Initial Data** *Initial Data Collection Report*	**Select Data** *Rationale for Inclusion/ Exclusion*	**Select Modeling Techniques** *Modeling Technique* *Modeling Assumptions*	**Evaluate Results** *Assessment of Data Mining Results w.r.t. Business Success Criteria* *Approved Models*	**Plan Deployment** *Deployment Plan*
Assess Situation *Inventory of Resources* *Requirements, Assumptions, and Constraints* *Risks and Contingencies* *Terminology* *Costs and Benefits*	**Describe Data** *Data Description Report*	**Clean Data** *Data Cleaning Report*	**Generate Test Design** *Test Design*	**Review Process** *Review of Process*	**Plan Monitoring and Maintenance** *Monitoring and Maintenance Plan*
Determine Data Mining Goals *Data Mining Goals* *Data Mining Success Criteria*	**Explore Data** *Data Exploration Report*	**Construct Data** *Derived Attributes* *Generated Records*	**Build Model** *Parameter Settings* *Models* *Model Descriptions*	**Determine Next Steps** *List of Possible Actions* *Decision*	**Produce Final Report** *Final Report* *Final Presentation*
Produce Project Plan *Project Plan* *Initial Assessment of Tools and Techniques*	**Verify Data Quality** *Data Quality Report*	**Integrate Data** *Merged Data*	**Assess Model** *Model Assessment* *Revised Parameter Settings*		**Review Project** *Experience Documentation*
		Format Data *Reformatted Data*			
		Dataset *Dataset Description*			

https://www.datascience-pm.com/crisp-dm-2/

Business Understanding



BACKGROUND:

With the rapid development of telecommunication industry, the service providers are inclined more towards expansion of the subscriber base. To meet the need of surviving in the competitive environment, the retention of existing customers has become a huge challenge. It is stated that the cost of acquiring a new customer is far more than that for retaining the existing one. Therefore, it is imperative for the telecom industries to use advanced analytics to understand consumer behavior and in-turn predict the association of the customers as whether or not they will leave the company.

Important points

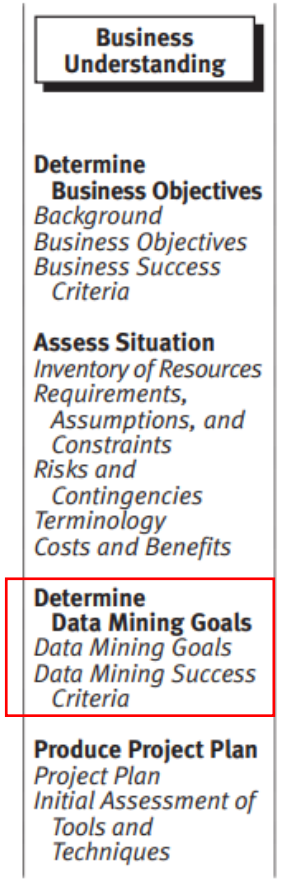
- Inclined customer
- Competitive environtment
- Cost acquiring customer > cost retaining customer
- Leaving company means revenue loss

Problem statement : how might we retain customer with high churn risk so we can prevent revenue loss ?

SOME BUSINESS OBJECTIVE AND SUCCES CRITERIA:

Business Objective	Success Criteria
Prevent revenue loss	Save at least at least 70% of revenue loss over the next period by retaining customers identified as high risk churned

Business Understanding



True	Churn Prediction (proba > threshold)	
	Low risk churn	High risk churn
Stay	Staying Customer	Inacurate Retention Effort
Churn	Undetected Churner	Detected Churner

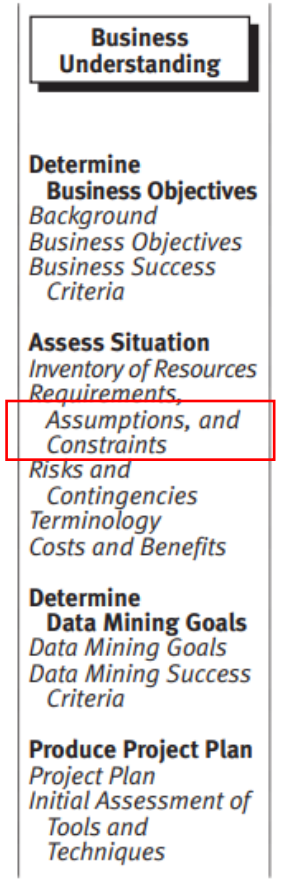
DATA MINING GOALS:

Minimize (Undetected Churner + Inacurate Retention Effort)

DATA MINING SUCCES CRITERIA:

We can minimize both “Undetected Churner” and “Inacurate Retention Effort” using classification metrics f1 score. By using this criteria in the end we hope we can prevent revenue loss as much as possible

Business Understanding



ASSUMPTION:

- retention cost take up 5% of monthly charges per customer
- let's say we are doing some retention campaign. succes rate is 80%.

Data Understanding

COLLECT INITIAL DATA:

In **data collection,** data was gathered from many sources such as

- Geographic
- Demographic
- Customer service usage
- Csutomer current profile

DESCRIBE DATA:

Some data description

Demographics

CustomerID: A unique ID that identifies each customer.

Count: A value used in reporting/dashboarding to sum up the number of customers in a filtered set.

Gender: The customer's gender: Male, Female

Age: The customer's current age, in years, at the time the fiscal quarter ended.

Senior Citizen: Indicates if the customer is 65 or older: Yes, No

Married: Indicates if the customer is married: Yes, No

Dependents: Indicates if the customer lives with any dependents: Yes, No. Dependents could be children, parents, grandparents, etc.

Number of Dependents: Indicates the number of dependents that live with the customer.

Services

CustomerID: A unique ID that identifies each customer.

Count: A value used in reporting/dashboarding to sum up the number of customers in a filtered set.

Quarter: The fiscal quarter that the data has been derived from (e.g. Q3).

Referred a Friend: Indicates if the customer has ever referred a friend or family member to this company: Yes, No

Number of Referrals: Indicates the number of referrals to date that the customer has made.

Tenure in Months: Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above.

Offer: Identifies the last marketing offer that the customer accepted, if applicable. Values include None, Offer A, Offer B, Offer C, Offer D, and Offer E.

Phone Service: Indicates if the customer subscribes to home phone service with the company: Yes, No

Avg Monthly Long Distance Charges: Indicates the customer's average long distance charges, calculated to the end of the quarter specified above.

Data Understanding

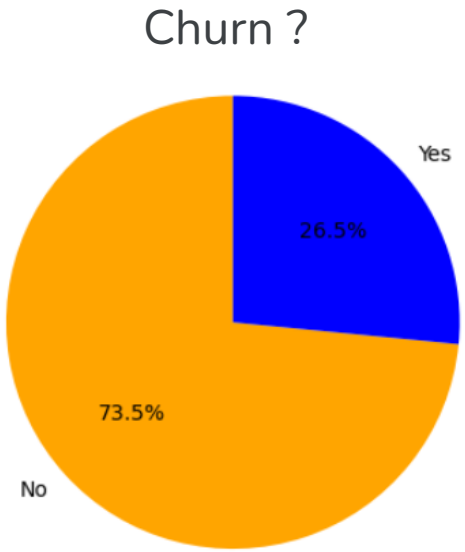
Data Understanding

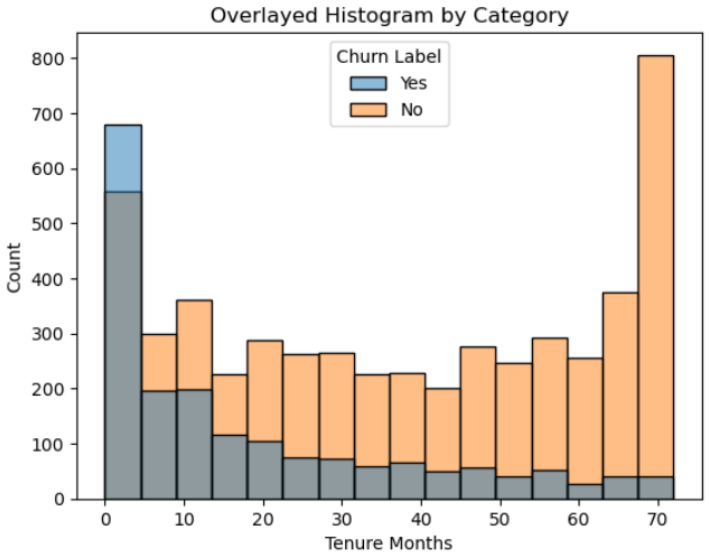
Collect Initial Data
Initial Data Collection Report

Describe Data
Data Description Report

Explore Data
Data Exploration Report

Verify Data Quality
Data Quality Report





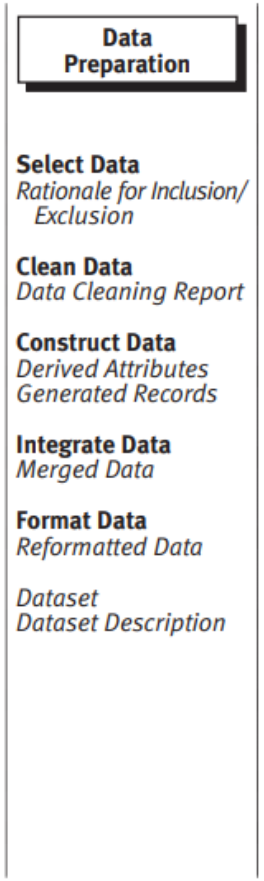
Explore Data:

- 26.5% customer labeled as churn. These customer can cause revenue loss for current period.
- So many factor related to churn. For example, old customer tend to stay than churn.

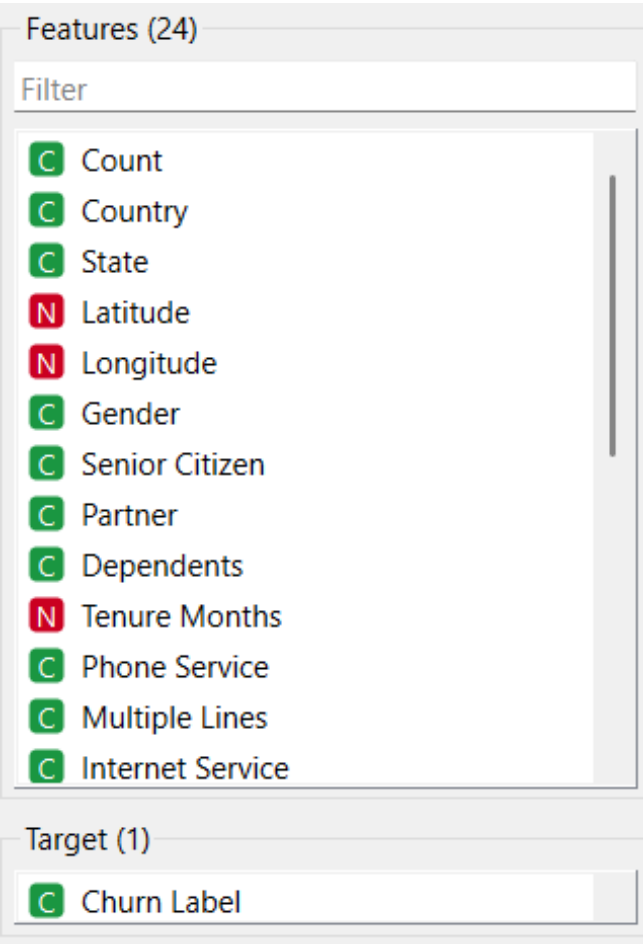
Monthly Charge Loss According to Current data:

- There are 7043 customer in current month : 1869 (26.5%) labeled as churn and 5174 (73.5%) labeled as non churn
- From 1869 churn customer, company will lose US$456,116 this month. We aim to decrease this loss.

Data Preparation



	Paperless Billing	Payment Method	Monthly Charges	Total Charges	Churn Label
1	Yes	Mailed check	53.85	108.15	Yes
2	Yes	Electronic check	70.70	151.65	Yes
3	Yes	Electronic check	99.65	820.50	Yes
4	Yes	Electronic check	104.80	3046.05	Yes
5	Yes	Bank transfer (a...	103.70	5036.30	Yes
6	No	Credit card (aut...	55.20	528.35	Yes
7	Yes	Electronic check	39.65	39.65	Yes
8	No	Mailed check	20.15	20.15	Yes
9	Yes	Electronic check	99.35	4749.15	Yes
10	No	Electronic check	30.20	30.20	Yes
11	Yes	Mailed check	64.70	1093.10	Yes
12	Yes	Electronic check	69.70	316.90	Yes
13	Yes	Electronic check	106.35	3549.25	Yes
14	Yes	Bank transfer (a...	97.85	1105.40	Yes
15	Yes	Electronic check	80.65	144.15	Yes
16	Yes	Credit card (aut...	99.10	1426.40	Yes
17	Yes	Credit card (aut...	80.65	633.30	Yes
18	Yes	Electronic check	95.45	1752.55	Yes



Modeling

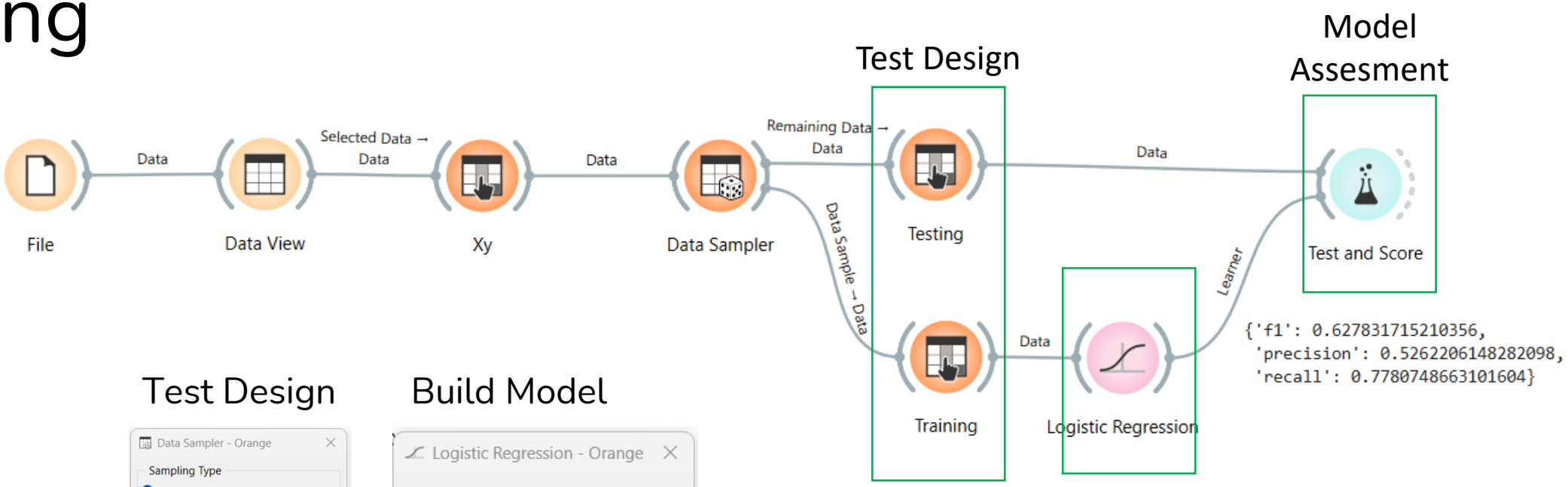
Modeling

Select Modeling Techniques
Modeling Technique
Modeling Assumptions

Generate Test Design
Test Design

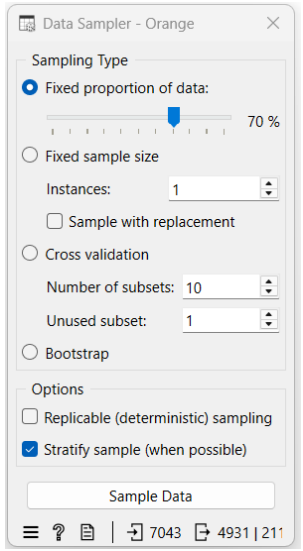
Build Model
Parameter Settings
Models
Model Descriptions

Assess Model
Model Assessment
Revised Parameter Settings

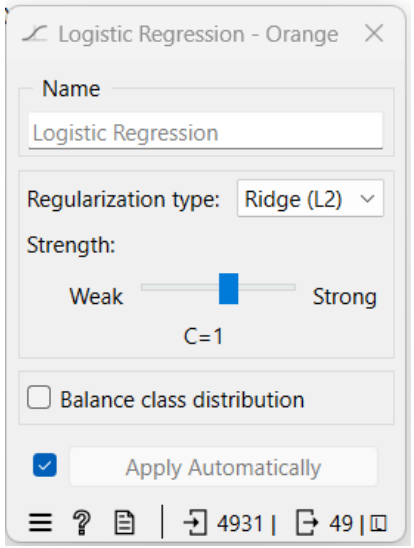


Modeling Technique

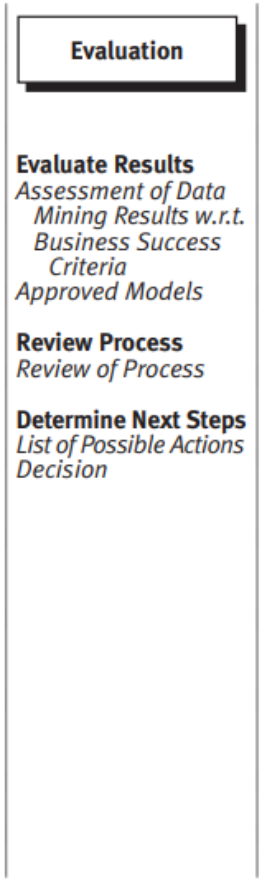
Test Design



Build Model



Evaluation



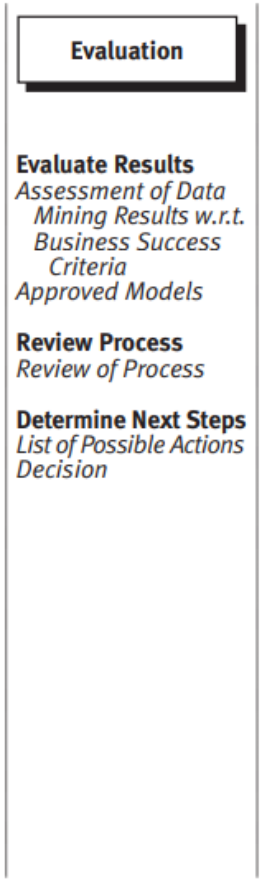
True	Churn Prediction (proba > threshold = 0.6)	
	Low Churn Risk	High Churn Risk
Stay	Staying Customer	Wasted Retention Cost
Churn	Undetected Monthly Charge Loss	Potentialy Saved Monthly Charge

Wasted Retention Cost = sum of (customer in “high risk” and staying customer) x (5% of monthly charge per customer)

Undetected Monthly Charge Loss = sum of (number of customer in “low risk” and churning customer) x monthly charge

Potentialy Saved Monthly Charge = sum of (number of customer in “high risk” and churning customer) x monthly charge per customer x retention campaign succes rate

Evaluation



EVALUATE RESULT:

"We can save up to 76.06% of potential monthly charge loss for the next period. For example, If potential loss US$112605 then we can save US$89442 (76.06% of potential loss)"

REVIE PROCESS:

It's still an MVP
ML model can still be optimized
Scenario could be different

DETERMINE NEXT STEP:

Model improvement (we can do some improvement next time!)
In real word scenario, Discuss more fitting scenario with business team

Evaluation

Evaluation

Evaluate Results
Assessment of Data Mining Results w.r.t. Business Success Criteria
Approved Models

Review Process
Review of Process

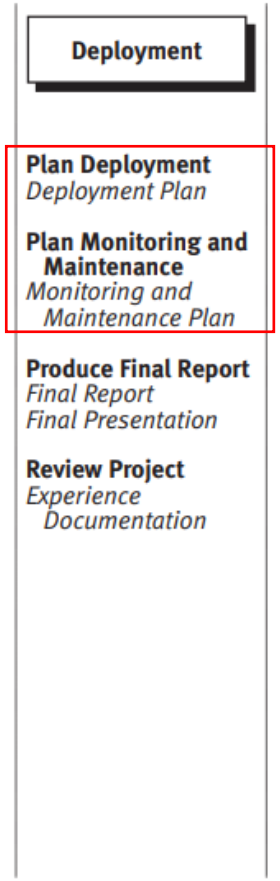
Determine Next Steps
List of Possible Actions
Decision

ANOTHER SCENARIO:

Another scenario that may be more fitting for business team

True	Predicted Churn Score		
	Low risk (score < threshold1)	Medium risk (threshold1 < score < threshold2)	High risk (score >threshold1)
Stay	Stay - low risk	Stay - medium risk	Stay - high risk
Churn	Churn - low risk	Churn - medium risk	Churn - high risk

Deployment



PLAN DEPLOYMENT:

- We can start from estimating revenue loss that can be saved every month
- Then adjust the budget for development cost-benefit wise.
- Whether it's cloud base or on premise

PLAN MONIORING AND MAINTENANCE:

- Data will be updated probably every month
- Data structure can be changed
- New variables probably coming
- Business requirement can be different and affect relevancy of the solution

Disclaimer

- I’m aware, ML model can still be optimized. It’s still an MVP!
- Scenario could be different
- Still so many thing can be improved. The take away is how to translate business problem into solution, measure it’s impact, and structurize the use case using CRIP-DM framework.
- We were trying to understand the use case and fit CRISP-DM using the available data. Some information may not be available.
- In reality, this will be far more complex. I’m trying to simplify it.