DATA SCIENCE PROJECT SCOPING WORKSHEET

1. Project Name: Health Insurance Cross Sell Prediction
2. Organization Name: Insurance All Company
3. Project Description: An insurance company that has provided Health Insurance to its customers now they need help in building a model to predict whether the policyholders (customers) from past year will also be interested in Vehicle Insurance provided by the company.
4. Who are the agencies/departments that will need to be involved? Data Science
5. Who are the individuals in these organizations that are stakeholders? What is their role? Insurance Health, selling insurance service.
6. GOALS (in order of priority)

What are you maximizing or minimizing?

Are there any constraints (budget, resources, etc.)?

|  |  |  |
| --- | --- | --- |
| Goal 1:  Costumers interested in buying Health Insurance with 20000 phone calls | Goal 2:  Increasing connections to 40,000.00 how much the profit will increase? (%) | Goal 3: |
| Constraint: Phone Calls | Constraint: | Constraint |

1. Actions

What is the action?

Who is taking the action?

What/Who is it being taken on?

How often?

|  |  |  |
| --- | --- | --- |
| Action 1: | Action 2: | Action 3: |
| Questions | Questions | Questions |
| a. What features we should target? | a. | a. |
| b. What’s the best model ROC score? | b. | b. |
| c. | c. | c. |
| d. | d. | d. |
|  |  |  |

1. Data
2. What Data do you have internally?

|  |  |  |
| --- | --- | --- |
| Data Source  Insurance Costumers perfil | Data Source | Data Source |
| What does it contain? Gender, Age, Driving License, Region Code, previously\_insured, vehicle\_age, vehicle\_damage, annual\_premium, policy\_sales\_channel, vintage, response | What does it contain? | What does it contain? |
| What level of granularity? Individual | What level of granularity? | What level of granularity? |
| How frequently is it collected/ updated? | How frequently is it collected/ updated? | How frequently is it collected/ updated? |
| Does it have unique identifiers that can be linked to other data source? 3 datasets: Train, test and ss | Does it have unique identifiers that can be linked to other data source? | Does it have unique identifiers that can be linked to other data source? |
| Other | Other | Other |

1. What data can you get externally and /or from public sources?

|  |  |  |
| --- | --- | --- |
| **Data Source** | **Data Source** | **Data Source** |
| What does it contain? | What does it contain? | What does it contain? |
| What level of granularity? | What level of granularity? | What level of granularity? |
| How frequently is it collected/ updated? | How frequently is it collected/ updated? | How frequently is it collected/ updated? |
| Does it have unique identifiers that can be linked to other data source? | Does it have unique identifiers that can be linked to other data source? | Does it have unique identifiers that can be linked to other data source? |
| Other | Other | Other |

1. What data would you need in addition to the ones above?

Data Source:

Data Source:

Data Source:

1. Analysis

What analysis needs to be done?

How will you validate the analysis?

|  |  |  |
| --- | --- | --- |
| Analysis 1:  Predicting if a costumer would be interested in Vehicle Insurance | Analysis 2: | Analysis 3: |
| Analysis type: Predcition | Analysis type: | Analysis type: |
| Which action will this analysis inform? Target people to buy the service | Which action will this analysis inform? | Which action will this analysis inform? |
| How will you validate this analysis? Use historical data to validate | How will you validate this analysis? | How will you validate this analysis? |