Data Analytics Lifecycle

Phase 1. Discovery

Initial Hypotheses are that GINA team thought this initiative would provide a means to share ideas globally and increase knowledge sharing among GINA members.

1. Learn the business domain
2. Farming the problem
3. Identifying key stakeholders
4. Interviewing the analytical sponsor
5. Developing initial Hypotheses
6. Identifying Potential Data Sources

Phase 2. Data Preparation

Capturing the results of its informal conversations with other leaders within its organisation, in academia, or in other organisations.

1. Prepare the analytical sandbox
2. Perform ETLT
3. Learn about the data
4. Data conditioning
5. Survey and visualise

Phase 3. Model Planning

Mine the data for patterns and insights to improve the team's operations and strategy.

1. Identifies candidate models to apply to data
2. Refers to the hypotheses developed in Phase 1.
3. Conduct Literature review of similar projects
4. Explore data and select variables

Phase 4. Model Building

Building a model of the information of the conversations.

1. Develop datasets for training, testing and production purposes
2. Train the analytical model and test it
3. Iterate back and forth for a while
4. Run models from software packages on file extracts and small datasets
5. Record the results and logic of the model during the phase
6. Record any operating assumptions made in the modelling process
7. Creating robust models to meet the objectives

Phase 5. Communicate Results

Evaluate and compare the results of built models in several aspects and make recommendations.

1. Compare the outcomes of the modelling to the criteria established for success and failure
2. Articulate the findings and outcomes to team members and stakeholders
3. Take into account caveats, assumptions, and any limitations of the results
4. Make recommendations for future work or improvements

Phase 6. Operationalize

Communicate the benefits of the project.

1. Communicate the benefits of the project more broadly
2. Set up a pilot project to deploy the work in a controlled way.
3. Manage risk
4. Learn the performance and constraints of the model.
5. Make adjustments before a full deployment.