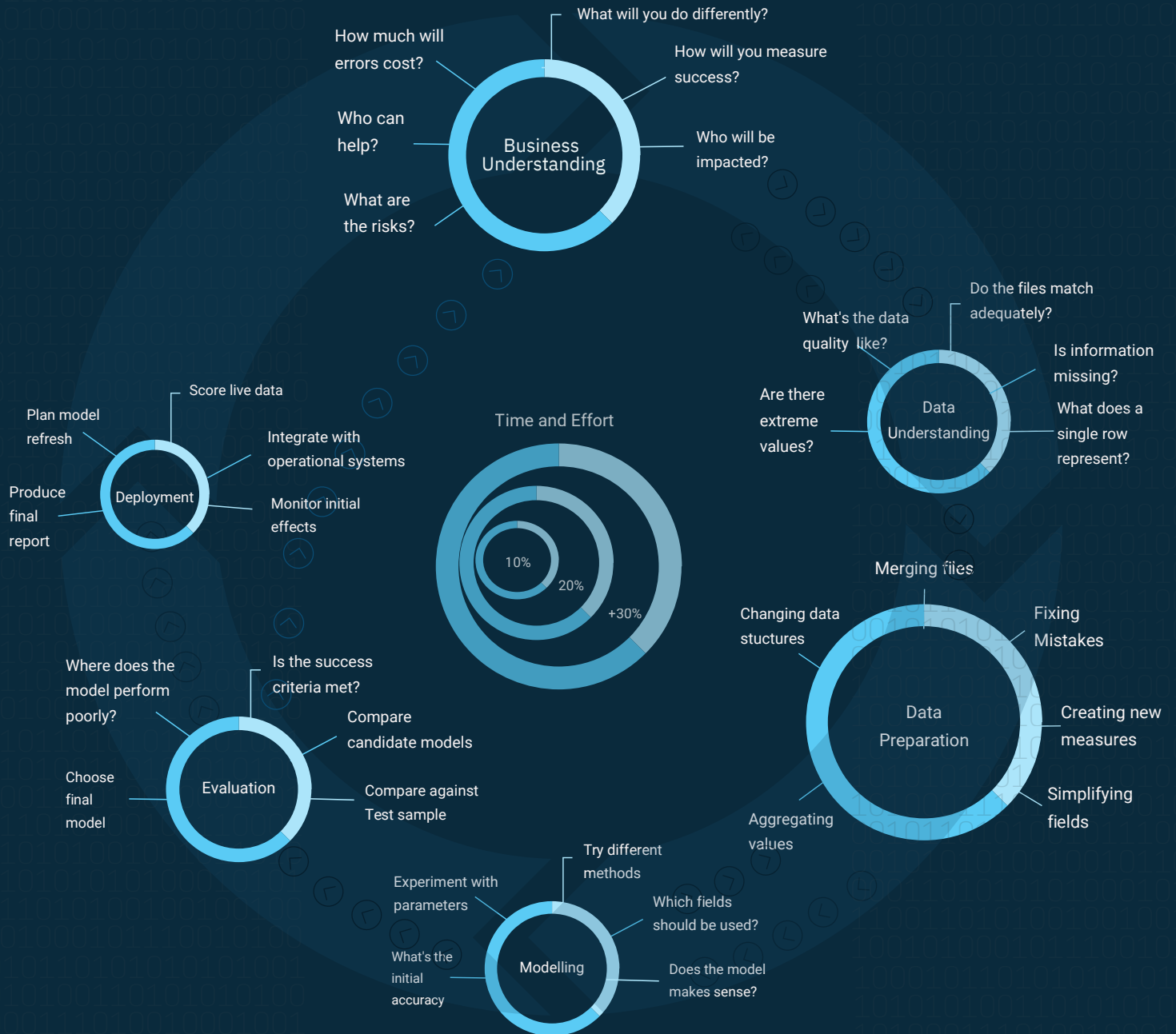# Planning Predictive Analytics Projects with CRISP-DM

The Cross Industry Standard Process for Data Mining

Why this subject matter, why not something else?

Is the outcome...
- Valuable?
- Measurable?
- Actionable?

Business Understanding
- What will you do differently?
- How much will errors cost?
- How will you measure success?
- Who can help?
- Who will be impacted?
- What are the risks?

Data Understanding
- What's the data quality like?
- Do the files match adequately?
- Are there extreme values?
- Is information missing?
- What does a single row represent?

Data Preparation
- Merging files
- Changing data stuctures
- Fixing Mistakes
- Creating new measures
- Simplifying fields
- Aggregating values

Modelling
- Try different methods
- Experiment with parameters
- Which fields should be used?
- What's the initial accuracy
- Does the model makes sense?

Evaluation
- Where does the model perform poorly?
- Is the success criteria met?
- Compare candidate models
- Choose final model
- Compare against Test sample

Deployment
- Score live data
- Plan model refresh
- Integrate with operational systems
- Produce final report
- Monitor initial effects

Time and Effort
- 10%
- 20%
- +30%

Products   Training   Events   Resources   Blog