

# Data Science Project Format

## The CRISP-DM Methodology

#	Stage	Considerations
1	<b>Business understanding</b>	<ul style="list-style-type: none"> <li>• <b>Problem statement:</b> Understand the context of the problem</li> <li>• <b>Objectives:</b> What is the purpose of the project?</li> <li>• <b>Needs:</b> What is the use case of the solution? Who are the users? Which features are necessary, and which are just nice to have?</li> <li>• <b>Data:</b> Identify data sources, an estimate of the data size and velocity, and data collection method.</li> <li>• <b>Definitions of success:</b> Determine the satisficing and optimising metrics, as well as the success and exit criteria.</li> <li>• <b>Presentation:</b> Dashboard, application, or report? Cloud or on-prem?</li> <li>• <b>Timelines:</b> Milestones</li> </ul>
2	<b>Data understanding</b> <i>Return to stage 1 if required</i>	<ul style="list-style-type: none"> <li>• <b>Data quality:</b> What are the quality issues and treatment options?</li> <li>• <b>EDA:</b> Inspect the distributions, statistical summaries, and units/scale. Explore whether there are any correlations or underlying patterns within the data.</li> </ul>
3	<b>Data preparation</b>	<ul style="list-style-type: none"> <li>• <b>Pre-processing:</b> Select and apply treatment to your data quality issues, do some feature engineering.</li> <li>• <b>ETL:</b> Integrate the data, perform transformations and scaling.</li> </ul>
4	<b>Modelling</b> <i>Return to stage 3 if required</i>	<ul style="list-style-type: none"> <li>• <b>Modelling:</b> Build a baseline model quickly, then iterate.</li> <li>• <b>Split:</b> Determine the train-dev-test split or consider cross-validation.</li> <li>• <b>Tuning:</b> Use the principles of orthogonalisation to guide you with hyperparameter tuning; consider grid search. Apply optimisation techniques if required.</li> <li>• <b>Comparison:</b> Use your metrics to guide whether your current idea is better than your previous idea.</li> <li>• <b>Iterate:</b> Iterate until either the success or exit criteria have been met.</li> </ul>
5	<b>Evaluation</b> <i>Return to stage 1 if required</i>	<ul style="list-style-type: none"> <li>• <b>Write-up:</b> Document the results of your trials as well as any reasoning for your decisions.</li> <li>• <b>Evaluation:</b> Ask the stakeholder to assess your final model against the business requirements.</li> </ul>
6	<b>Deployment</b>	<ul style="list-style-type: none"> <li>• <b>Plan:</b> Create a deployment plan and checklist.</li> <li>• <b>Clean up:</b> Convert any notebooks into scripts, draw diagrams for each script.</li> <li>• <b>Deploy:</b> Deploy your project, provide access to the stakeholders, and monitor the results.</li> </ul>