**MA01 Group Project – ACTION PLAN Strict Deadline & Flexible Task Plan**

**Final Project Deadline:** Monday, 10 November 2025

**Team Roles**

| **Member** | **Strengths** | **Responsibilities** |
| --- | --- | --- |
| **Erwin** | Coding, overview, planning | Lead project integration, pipelines, deployment, oversee progress |
| **Joel** | Coding | Implement ML models, feature engineering, testing, model optimization |
| **Tijn** | Coding & business writing | Support preprocessing, simple models, document tasks, contribute to slides |
| **Eva** | Business writing & coding support | Assist with visualizations, helper functions, document notebooks, prepare presentation |

**Note:** Tasks are **flexible**. Members are encouraged to contribute across areas they are interested in (e.g., modeling, evaluation, deployment), not strictly limited to assigned responsibilities.

**Strict Deadlines & Tasks**

**Phase 1: Initial Setup & Business Understanding**

**Deadline: Fri 18 Oct 2025**

* Setup repo, folders, main overview notebook
* Install dependencies, skeleton code for data loading
* Draft **Business Understanding notebook**, define ML target & success criteria
* Draft EDA outline, placeholders for charts

**Deliverables:** Repo ready, Business Understanding notebook, EDA outline

**Phase 2: Data Understanding**

**Deadline: Fri 25 Oct 2025**

* Inspect dataset, missing values, data types, initial cleaning
* Implement preprocessing functions (encoding, normalization)
* Generate visualizations, descriptive stats tables
* Summarize dataset insights in text and visuals

**Deliverables:** Cleaned subset of dataset, EDA notebook, summary tables

**Phase 3: Data Preparation & Specialization Topic**

**Deadline: Wed 30 Oct 2025**

* Implement preprocessing pipeline and feature engineering
* Research and document specialization topic (dimensionality reduction, data cleaning, or outlier detection)
* Assist with preprocessing, generate visualizations, prepare interim presentation content

**Deliverables:** Fully preprocessed dataset, interim presentation content, updated notebooks

**Phase 4: Modeling**

**Deadline: Mon 4 Nov 2025**

* Build pipeline and composite estimator, integrate preprocessing + models
* Train models (Random Forest, Gradient Boosting, Linear Regression) and collect metrics
* Implement additional models (e.g., Linear Regression, KNN) and integrate into pipeline
* Generate result plots and summarize model performance

**Deliverables:** Trained models, evaluation tables/plots, modeling notebook

**Flexibility:** Any member interested in modeling can contribute or take lead on specific models or experiments.

**Phase 5: Evaluation & Optimization**

**Deadline: Thu 7 Nov 2025**

* Hyperparameter tuning, cross-validation, SHAP/LIME explainability
* Generate metrics tables, create visualizations
* Update report & slides with evaluation insights

**Deliverables:** Optimized models, evaluation charts, business-focused summaries

**Flexibility:** Members can assist in evaluation, visualization, or adding explainable AI elements.

**Phase 6: Deployment & Presentation**

**Deadline: Sun 9 Nov 2025**

* Finalize end-to-end pipeline, demo notebook or Streamlit app
* Test pipeline and ensure reproducibility
* Integrate models into demo, refine slides and visuals
* Document demo and report content

**Deliverables:** Final pipeline/demo, slides, report

**Flexibility:** Team members can help with deployment tasks, presentations, or demo design.

**Phase 7: Submission**

**Deadline: Mon 10 Nov 2025**

* Submit notebooks, pipeline, report, presentation
* Verify all files are organized and documented in repo

**Optional Features to Impress Lecturers**

1. Deployable pipeline that accepts new rider data
2. Explainable AI elements (SHAP/LIME)
3. Interactive visualizations or dashboards
4. Advanced evaluation metrics for robustness
5. Experiment with ensembles or stacking

**Key Notes**

* Work is **distributed but flexible**: members can contribute to any task they are interested in.
* Main overview notebook should **link to all step notebooks** for easy navigation.
* Strict deadlines provide buffer days before submission.
* Daily check-ins recommended **from Phase 3 onward**.