# Assignment Coverage Matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Task | Time Taken | Approach Used | Issues Faced | Status |
| Data Preparation (Users & Events) | 1.5 hrs | Loaded Parquet files using PyArrow; structured into Pandas DataFrames for processing. | File path adjustments, ensuring large file handling without memory overload. | Completed |
| Customer-Journey Graph in Neo4j | 2 hrs | Created (:Cluster) and (:Customer) nodes; linked via [:IN\_CLUSTER]; excluded raw events. | Initial schema mismatch; adjusted to store only required attributes. | Completed |
| Clustering Logic | 2 hrs | Used grouping & simple heuristic rules to form clusters based on device/category/event patterns. | Balancing cluster size; ensuring representational summaries. | Completed |
| GraphRAG Retrieval | 2.5 hrs | Implemented graph filters + FAISS semantic ranking on cluster summaries. | Neo4j query errors due to missing properties; resolved by schema alignment. | Completed |
| Ask & Show Agent | 2 hrs | Rule-based NLP parsing for device/category/funnel flags; integrated with GraphRAG retrieval. | Keyword coverage tuning; avoiding overfitting to example queries. | Completed |
| Testing & Example Queries | 1 hr | Ran multiple example queries to verify ranking & filtering logic. | Some queries returned empty due to sparse clusters; resolved by adjusting filters. | Completed |
| Documentation | 1 hr | Created README, coverage matrix, and .docx project documentation. | Ensuring clarity & aligning with assignment deliverables. | Completed |