Dataset Options Guide (Capstone Alternatives)

This guide outlines how students can select an approved alternative dataset for their **capstone project only, by approved exception**. All weekly projects must use the absolute default REIT Master dataset (monthly, REIT-level).

Approved Alternative Sources

Orbis Open Dataset Catalog

- Browse and select datasets from the **Orbis Open Dataset Catalog** at phdai.ai
- The curated catalog is also available as `OpenData_rows.csv`
- Best for: student-selected topics (finance, labor, climate, crypto, etc.)
- You can choose any dataset from the Orbis open dataset catalog that meets the minimum requirements below

---

Minimum Requirements (Alternatives)

Your dataset must meet these minimums:

- **Entities:** At least 50 entities (e.g., firms, regions, countries, cryptocurrencies)
- **Time span:** At least 5 years of data for time series or panel data
- **Observations:** Sufficient observations to support analysis (typically 500+ rows for cross-sectional data, or 50+ entities × 60+ time periods for panel data)
- **Variables:** At least 1 outcome variable + explanatory variables
- **Documentation:** Clear source link and access method
- **Reproducibility:** Data access and cleaning steps must be scriptable

Ideal dataset: 10 years of data provides more robust analysis and stronger statistical power, but 5 years is the minimum.

If you cannot meet these requirements, you must use the default track (REIT Master dataset, monthly REIT-level).

Proposal (Due Week 4, Friday)

Submit a 1-page proposal that includes:

- **Dataset source:** URL and access steps
- **Unit of analysis:** Entity and time unit (e.g., firm-month, region-quarter)
- **Research question:** Clear, testable question
- **Key variables:** Outcome + core predictors
- **Feasibility:** Missingness, access limits, and cleaning steps

Proposals are approved within one week. If not approved, you will default to the REIT Master dataset track.

Mapping to Capstone Milestones

Alternative datasets follow the same milestones as the default track:

- **Milestone 1:** Data pipeline and clean panel/time series
- **Milestone 2:** EDA dashboard and hypothesis development
- **Milestone 3:** Econometric models (FE, DiD, or justified alternative)
- **Milestone 4:** Final memo + presentation

Examples (Acceptable Proposals)

Example 1: Orbis Open Dataset Catalog

- **Question:** Do unemployment shocks predict rental price changes across metros?
- **Entity/Time:** Metro-month
- **Outcome:** Rent index growth
- **Predictors:** Unemployment rate, wage growth, migration inflow
- **Source:** Selected from Orbis Open Dataset Catalog at phdai.ai

Example 2: Orbis Open Dataset Catalog

- **Question:** How do interest rate changes affect cryptocurrency market volatility?
- **Entity/Time:** Cryptocurrency-month
- **Outcome:** Volatility measure (e.g., realized volatility)
- **Predictors:** Federal funds rate, market cap, trading volume
- **Source:** Selected from Orbis Open Dataset Catalog at phdai.ai

---

Approval Criteria

Instructors will check:

- Data meets minimum requirements
- Research question is feasible in 15 weeks
- Variables support causal or predictive modeling
- Data access and cleaning are reproducible

Submission Notes

- Include an **AI Audit Appendix** if AI tools helped you choose or scope the dataset
- If your dataset access changes mid-semester, notify the instructor immediately