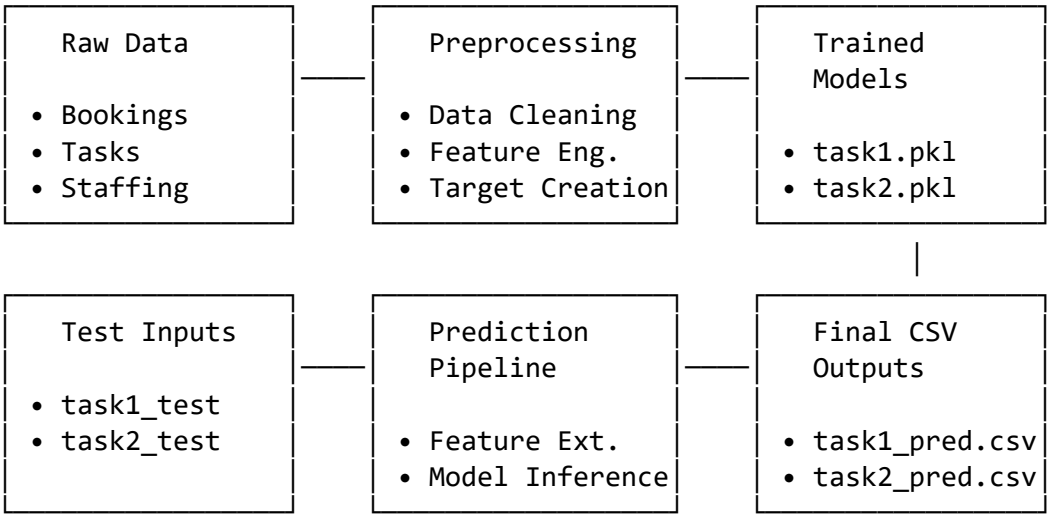
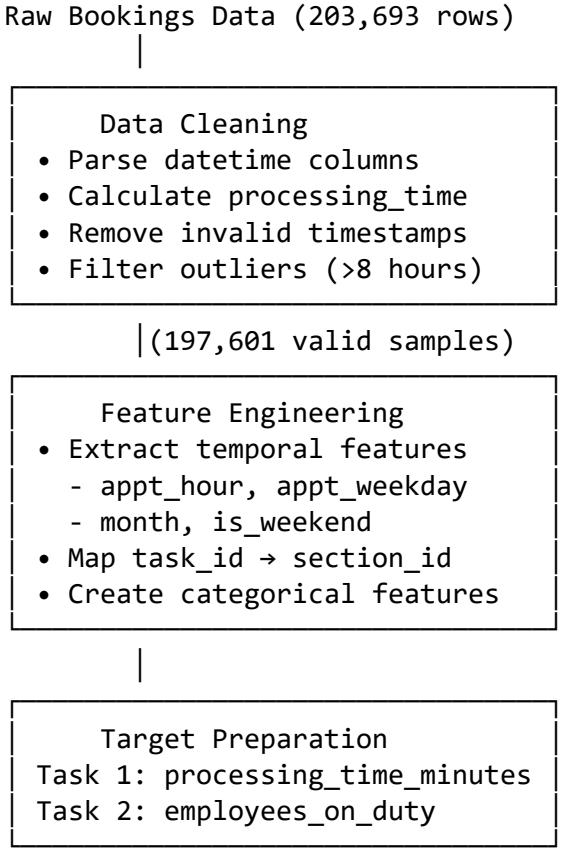
Architecture Diagrams

BitByBit Datathon 2025 - Team BitByBit

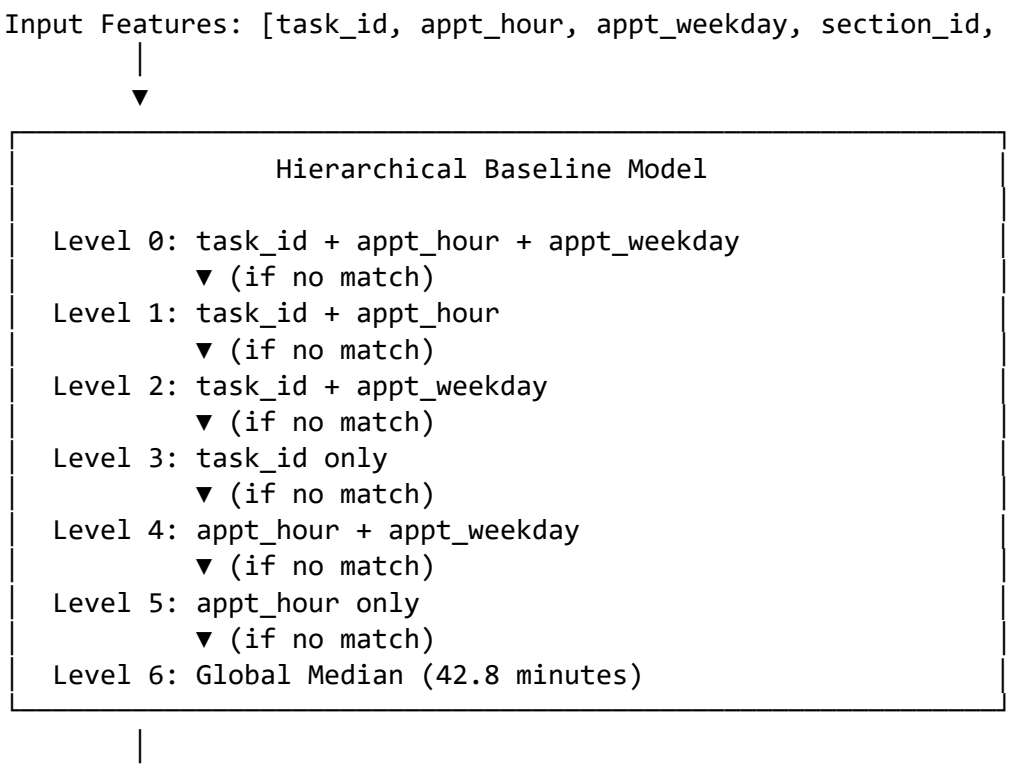
1. Overall System Architecture



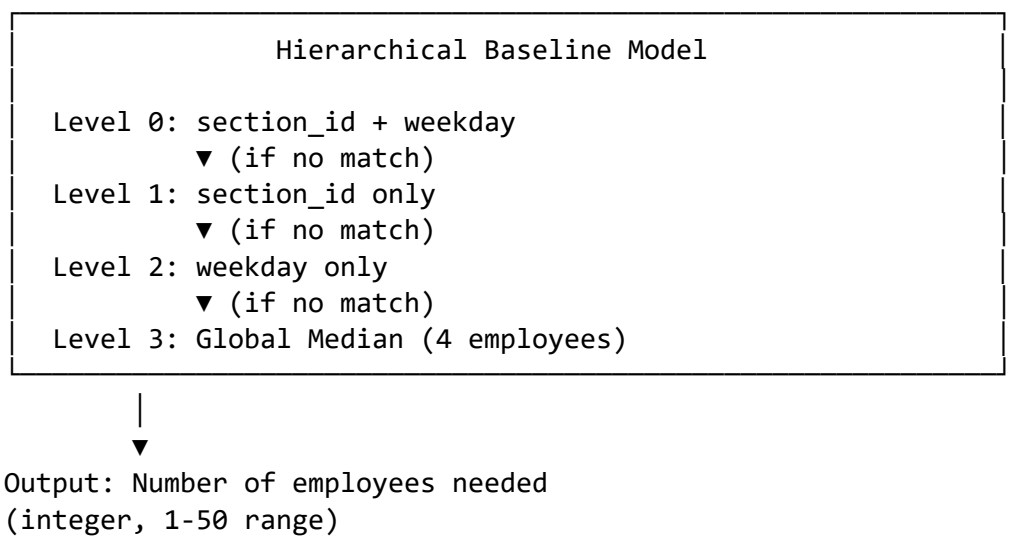
2. Data Preprocessing Pipeline



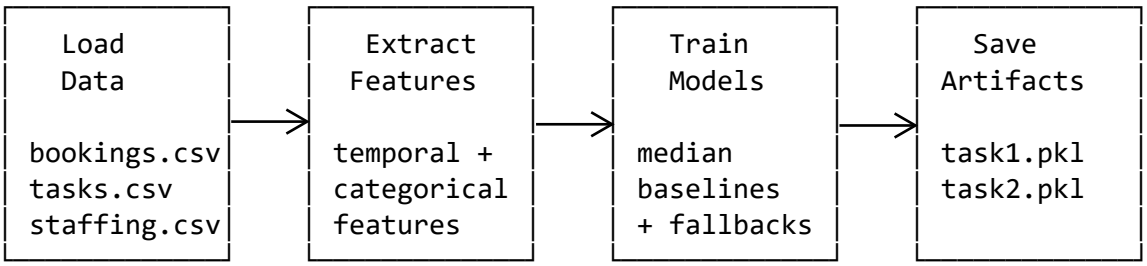
3. Model Architecture - Task 1 (Processing Time)



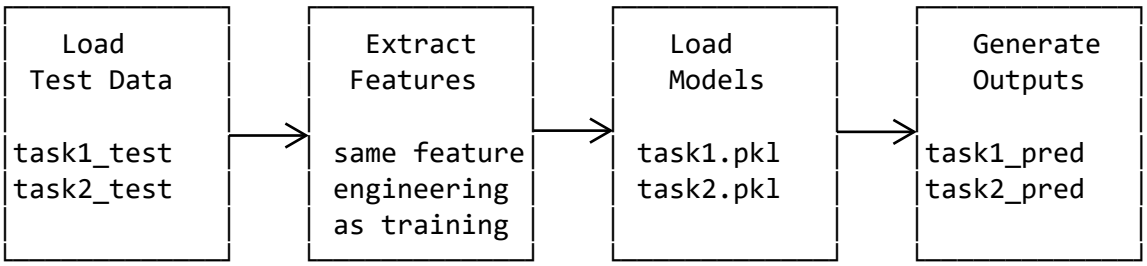
4. Model Architecture - Task 2 (Staffing)



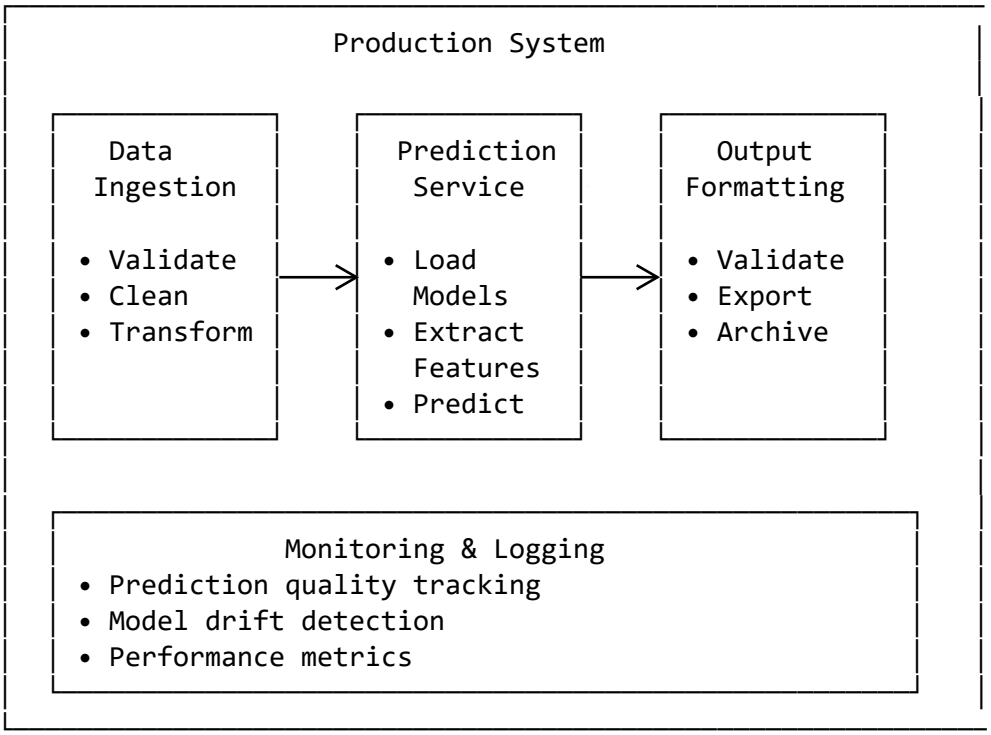
5. Training Pipeline Flow



6. Prediction Pipeline Flow



7. Deployment Architecture (High Level)



8. Technology Stack

Tech Stack	
Data Processing:	pandas, numpy
Modeling:	scikit-learn (optional), custom baselines
Validation:	Custom validation functions
I/O:	CSV, pickle serialization
Logging:	Python logging module
Environment:	Python 3.8+, cross-platform

9. Key Design Decisions

- **Baseline-First Approach**: Prioritized robust statistical methods over complex ML
- **Hierarchical Fallbacks**: Ensured 100% prediction coverage for all test cases
- **Modular Architecture**: Separated preprocessing, training, and prediction concerns
- **Defensive Programming**: Comprehensive error handling and data validation
- **Business Logic Integration**: Applied domain knowledge (working hours, staffing limits)

10. Performance Characteristics

Training Performance:

Task	MAE	RMSE	Samples
Task 1	14.37 min	19.74 min	197,601
Task 2	0.60 emp	0.80 emp	5,802

Prediction Coverage: 100% (no failed predictions)
Model Size: <1MB per model (lightweight .pkl files)
Execution Time: <30 seconds for full pipeline