**Requirements Specifications**

**P13: ContinuumAI**

|  |  |
| --- | --- |
| **Student ID** | **Name** |
| **26100059** | **Ali Faizan Chughtai** |
| **26100029** | **Muhammad Nafees** |
| **26100038** | **Muhammad Mustufa** |
| **26100146** | **Muhammad Bazaf Shakeel** |
| **26100063** | **Umer Raja** |

|  |  |  |
| --- | --- | --- |
| **Content** | **Totals** | **Obtained** |
| **Introduction & system actors** | **5** | **5** |
| **Use case diagram** | **20** | **10** |
| **Use case descriptions** | **20** | **16** |
| **Class diagram** | **20** | **18** |
| **Sequence diagram** | **20** | **16** |
| **State diagram** | **5** | **5** |
| **Data Requirements** | **20** | **20** |
| **Non-functional requirements** | **10** | **10** |
| **Security requirements** | **10** | **10** |
| **Use of generative AI** | **5** | **5** |
| **Who did what** | **5** | **5** |
| **Review checklist** | **5** | **5** |
| **Overall formatting/template** | **5** | **4** |
| **Late submission penalty** | **-20** | **-** |
| **GitHub Folder structure penalty** | **-5** | **-** |
| **Grand Total** | **150** | **129** |
| **General Comments/Individual Grading:** | | |

**Table of Contents**

[1.](#_laj7qto2qtth) Introduction 3

[2.](#_z8ts9sahj983) System Actors 4

[3.](#_tz91d98bagko) Use Cases 5

[3.1 Use Case Diagrams 5](#_82bgsh40qy7i)

[3.2 Description of Use Cases 5](#_acitucm8iywq)

[3.2.1 Withdraw cash 5](#_yd8jcx5nc17o)

[3.2.2 Transfer funds 6](#_q3dlqymplo5l)

[4.](#_86zgrd4urt7b) Class Diagram 7

[4.1 Diagram 7](#_rrlju2j8rx94)

[4.2 Description 7](#_6tivneb97f26)

[5.](#_wiknu6sngyo8) Sequence Diagrams 8

[5.1 Use case Name e.g., Withdraw cash 8](#_o0jiqvp58zu9)

[5.2 Use case Name e.g., Transfer funds 8](#_r976pmwf0xov)

[6.](#_clypa9in5n5o) State Diagrams 9

[6.1 Diagram details 9](#_oweko5b31k6)

[6.2 Diagram 9](#_230223bpwa7d)

[7.](#_63gvslq2ba3i) Data Requirements 10

[8.](#_lnlvfxjuydli) Non-functional Requirements / Quality Attributes 11

[9.](#_81191w8avkfs) Security Requirements 12

[10.](#_miml5va8d9g2) Security Engineer 13

[11.](#_xew7xd3hf0tp) Use of Generative AI 14

[12.](#_uunbai8t7mzv) Who Did What? 15

[13.](#_jou2y0btcdo6) Review checklist 15

# Introduction

**Project Overview:** ContinuumAI is an Agentic AI system designed to function as a personal data scientist for non-technical users in business roles. Its primary objective is to enable decision-makers - such as managers in HR, Sales, Marketing, and Product teams - to access and act upon data insights through natural language prompts, without requiring coding or technical expertise.

The system bridges the gap between data science capabilities and business needs by automating the full analytics workflow: from data ingestion to descriptive, diagnostic, predictive, and prescriptive analysis. It provides outputs such as trends, visualizations (e.g., bar charts, line graphs), forecasts, performance metrics, and actionable recommendations entirely based on user queries in plain English.

For example, a regional sales manager might want to understand what factors influenced last quarter’s revenue dip or forecast next month's performance. Instead of relying on a data team, they can simply ask the system in natural language. ContinuumAI automatically identifies the relevant data, applies the right analytical methods, and delivers clear, actionable insights empowering the user to make informed decisions instantly.

While the long-term goal is to support **all major business domains**, the current version of ContinuumAI is focused specifically on the **Sales domain**, allowing us to build deep and meaningful capabilities before expanding further. (Hence, for this document as well we will be focussing on sales only)

As for the data required to support our sales-oriented features, we are currently in discussions with firms to obtain sales related datasets, once a functional MVP has been achieved. For initial development, we will explore open-source alternatives (e.g., Kaggle, Hugging Face) to develop an idea of standard dataset forms and generate synthetic data based on well-defined schemas informed by the open-set datasets as well as input from industry advisors, in light of the use cases detailed in this document.

# System Actors

|  |  |
| --- | --- |
| **Actor Name** | **Description** |
| Sales Manager | Responsible for overseeing sales teams and performance across regions or product lines. Uses ContinuumAI to track sales trends, identify underperforming areas, and generate forecasts or performance reports. |
| Product Manager | Drives product development and strategy. Uses the system to uncover sales trends, customer insights, and market opportunities, enabling faster, data-informed decisions without technical support. |
| Sales Business Analyst | Supports sales strategy through data analysis and reporting. Uses the platform to quickly explore datasets, validate hypotheses, and generate insights (e.g., customer churn, revenue patterns) without manual SQL or spreadsheet work. |
| Customer Relationship Manager | Works post-sale to ensure client satisfaction and retention. Uses the system to identify at-risk customers based on declining purchases or engagement trends and to suggest proactive follow-up actions. |
| Regional Sales Lead | Focuses on sales activities within a specific geographic region. Uses ContinuumAI to assess territory-specific trends, seasonal patterns, and demand shifts, helping plan outreach and future steps to take. |
| ERP/Resource Planning Manager | Responsible for aligning inventory, procurement, and logistics based on product demand. Uses ContinuumAI to monitor sales trends across regions and time periods, helping to allocate resources (e.g., stock, labor, logistics support) more efficiently toward high-performing or fast-moving products. |
| Marketing Strategist | Focuses on planning high-level marketing campaigns and aligning them with business goals. Uses ContinuumAI to analyze sales trends, customer demographics, and regional performance data to identify underperforming products or markets. Based on these insights, they tailor marketing efforts to boost visibility and engagement where it's most needed. |

# Use Cases

## 

## 

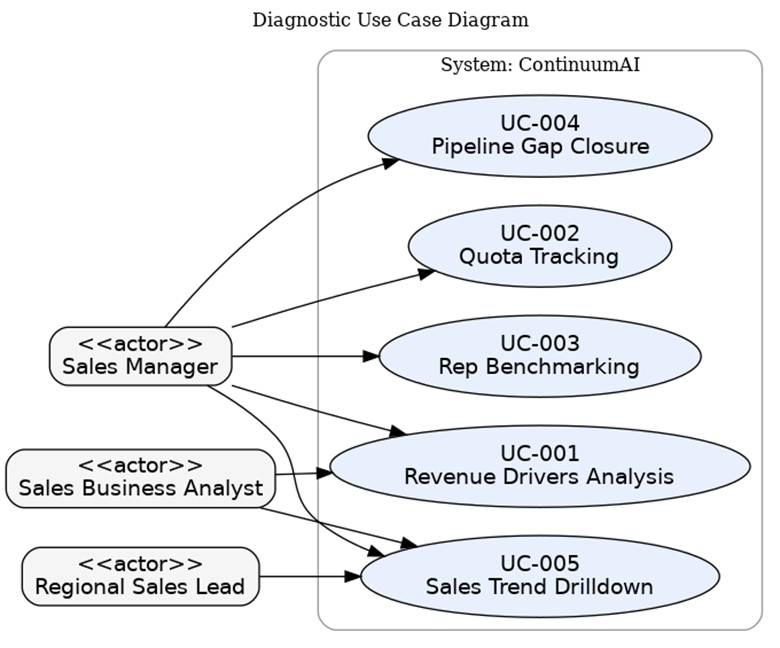
## Use Case Diagrams

Update this section, use standard UML notation for use case diagram

Use the same format as given in the template document for use case description.

The following figures show grouped Use Case Diagrams for the ContinuumAI Sales MVP. Each diagram highlights actors and their interactions with use cases in a specific analytical domain (Diagnostic, Predictive, Prescriptive, and Operational/Customer).

## Diagnostic Domain

****

## Predictive Domain

## Prescriptive DomainOperational/Customer Domain

## Description of Use Cases

### 

## UC-001 Revenue Drivers Analysis

|  |  |
| --- | --- |
| **Identifier** | UC-001 |
| **Purpose** | Identify the top factors contributing to a quarterly revenue dip. |
| **Pre-conditions** | * Historical sales data for the last quarter is available. * The Sales Manager is logged in. |
| **Post-conditions** | A diagnostic report is generated showing top 3 drivers of revenue decline. |
| **Typical Course of Action** | 1. Sales Manager enters query: 'What caused the revenue dip last quarter?'  2. System fetches relevant KPIs (sales by region, product, rep, discounting).  3. System applies root-cause analysis.  4. System generates report with charts and plain-English summary.  5. Sales Manager views or exports report. |
| **Alternate Courses of Action** | Manager selects predefined 'Revenue Drivers' template instead of typing query. |
| **Exception Paths** | If data incomplete, error message: 'Insufficient data available for requested period.' |

## UC-002 Quota Tracking

|  |  |
| --- | --- |
| **Identifier** | UC-002 |
| **Purpose** | Monitor real-time quota attainment. |
| **Pre-conditions** | * Sales quota targets uploaded. * Live sales data ingested. |
| **Post-conditions** | Dashboard displays % attainment vs. target. |
| **Typical Course of Action** | 1. Manager requests 'Show quota attainment by rep.' 2. System compares targets vs. actuals. 3. Dashboard shows attainment and highlights risks. |
| **Alternate Courses of Action** | Manager requests drilldown by region instead of rep. |
| **Exception Paths** | Missing targets → system displays 'Quota data not found.' |

## UC-003 Rep Benchmarking

|  |  |
| --- | --- |
| **Identifier** | UC-003 |
| **Purpose** | Compare sales rep performance on defined KPIs. |
| **Pre-conditions** | Historical rep-level sales data exists. |
| **Post-conditions** | Comparison chart generated. |
| **Typical Course of Action** | 1. Manager asks: 'Compare reps by average deal size and win rate.' 2. System retrieves and aggregates rep data. 3. System displays comparative table and visuals. |
| **Alternate Courses of Action** | Sort reps by revenue contribution. |
| **Exception Paths** | If only one rep has data → show 'Not enough data for comparison.' |

## UC-004 Pipeline Gap Closure

|  |  |
| --- | --- |
| **Identifier** | UC-004 |
| **Purpose** | Recommend actions to close pipeline shortfall. |
| **Pre-conditions** | CRM pipeline data ingested. |
| **Post-conditions** | List of recommended actions provided. |
| **Typical Course of Action** | 1. Manager requests: 'How do I close my pipeline gap this quarter?'  2. System evaluates pipeline vs. target.  3. Recommendations generated (e.g., increase calls, upsell). |
| **Alternate Courses of Action** | User simulates different strategies. |
| **Exception Paths** | Missing pipeline data → error message. |

## UC-005 Sales Trend Drilldown

|  |  |
| --- | --- |
| **Identifier** | UC-005 |
| **Purpose** | Analyze sales trends by region/product. |
| **Pre-conditions** | Time-series sales data available. |
| **Post-conditions** | Trend visual shown. |
| **Typical Course of Action** | 1. User enters query: 'Show sales trend for Product A in North region.' 2. System filters dataset. 3. Line chart displayed with summary. |
| **Alternate Courses of Action** | User drills down further by customer segment. |
| **Exception Paths** | No sales for product → 'No data available.' |

## UC-006 Territory Forecasting

|  |  |
| --- | --- |
| **Identifier** | UC-006 |
| **Purpose** | Predict territory sales for next quarter. |
| **Pre-conditions** | Regional sales history available. |
| **Post-conditions** | Forecast chart produced. |
| **Typical Course of Action** | 1. Regional Lead requests: 'Forecast next quarter’s sales in East.' 2. System applies forecasting model. 3. Predicted values displayed with confidence intervals. |
| **Alternate Courses of Action** | Forecast by month instead of quarter. |
| **Exception Paths** | Too little history → display 'Forecast not possible.' |

## UC-007 What-If Pricing Simulation

|  |  |
| --- | --- |
| **Identifier** | UC-007 |
| **Purpose** | Simulate impact of price/discount changes. |
| **Pre-conditions** | Pricing and sales elasticity data available. |
| **Post-conditions** | Simulated revenue impact displayed. |
| **Typical Course of Action** | 1. Manager asks: 'What if discounts increase by 10%?' 2. System models impact using historical elasticity. 3. Report generated with predicted revenue change. |
| **Alternate Courses of Action** | Test multiple discount levels. |
| **Exception Paths** | Elasticity data missing. |

## UC-008 Product Demand Forecast

|  |  |
| --- | --- |
| **Identifier** | UC-008 |
| **Purpose** | Forecast SKU-level demand. |
| **Pre-conditions** | SKU-level sales history ingested. |
| **Post-conditions** | Forecasted demand values displayed. |
| **Typical Course of Action** | 1. ERP Manager asks: 'Forecast demand for SKU-101 next month.' 2. System applies forecasting algorithm. 3. Demand prediction shown. |
| **Alternate Courses of Action** | Aggregate forecast for entire category. |
| **Exception Paths** | SKU not found. |

## UC-009 Campaign Impact Forecast

|  |  |
| --- | --- |
| **Identifier** | UC-009 |
| **Purpose** | Predict sales uplift from campaigns. |
| **Pre-conditions** | Campaign and sales data linked. |
| **Post-conditions** | Forecast chart shown. |
| **Typical Course of Action** | 1. Marketing Strategist requests: 'Forecast impact of summer campaign on sales.' 2. System links campaign spend and historical response. 3. Uplift forecast shown. |
| **Alternate Courses of Action** | Forecast by channel. |
| **Exception Paths** | No campaign data. |

## UC-010 Churn Risk Forecasting

|  |  |
| --- | --- |
| **Identifier** | UC-010 |
| **Purpose** | Predict churn risk segments. |
| **Pre-conditions** | Customer transaction history available. |
| **Post-conditions** | Churn risk scores displayed. |
| **Typical Course of Action** | 1. CRM asks: 'Which customers are at risk of churn?' 2. System applies churn prediction model. 3. At-risk customers flagged. |
| **Alternate Courses of Action** | Filter by revenue tier. |
| **Exception Paths** | Missing customer data. |

## UC-011 At-Risk Accounts

|  |  |
| --- | --- |
| **Identifier** | UC-011 |
| **Purpose** | Detect accounts with declining purchases. |
| **Pre-conditions** | Account purchase history exists. |
| **Post-conditions** | At-risk accounts flagged. |
| **Typical Course of Action** | 1. CRM requests declining accounts report. 2. System analyzes purchase frequency. 3. Accounts with >30% drop highlighted. |
| **Alternate Courses of Action** | View trend by region. |
| **Exception Paths** | No accounts meet criteria. |

## UC-012 Upsell/Cross-Sell Suggestions

|  |  |
| --- | --- |
| **Identifier** | UC-012 |
| **Purpose** | Recommend additional products. |
| **Pre-conditions** | Customer purchase histories available. |
| **Post-conditions** | Recommendation list generated. |
| **Typical Course of Action** | 1. CRM asks: 'Suggest cross-sell for Account A.' 2. System analyzes purchase patterns. 3. Recommendation list provided. |
| **Alternate Courses of Action** | Recommendations for all accounts. |
| **Exception Paths** | Insufficient purchase history. |

## UC-013 Customer Health Dashboard

|  |  |
| --- | --- |
| **Identifier** | UC-013 |
| **Purpose** | View overall customer health indicators. |
| **Pre-conditions** | Engagement and satisfaction data available. |
| **Post-conditions** | Health dashboard displayed. |
| **Typical Course of Action** | 1. CRM requests: 'Show health of Account B.' 2. System aggregates satisfaction, purchase, usage data. 3. Health score displayed. |
| **Alternate Courses of Action** | Export dashboard to PDF. |
| **Exception Paths** | Missing engagement data. |

## UC-014 Renewal Likelihood Prediction

|  |  |
| --- | --- |
| **Identifier** | UC-014 |
| **Purpose** | Predict contract renewal probability. |
| **Pre-conditions** | Contract history ingested. |
| **Post-conditions** | Renewal probability displayed. |
| **Typical Course of Action** | 1. CRM asks: 'Likelihood of renewal for Account C?' 2. System applies prediction model. 3. Renewal probability shown. |
| **Alternate Courses of Action** | View renewal risk by customer tier. |
| **Exception Paths** | Contract data missing. |

## UC-015 Customer Cohort Analysis

|  |  |
| --- | --- |
| **Identifier** | UC-015 |
| **Purpose** | Compare performance of cohorts. |
| **Pre-conditions** | Cohort data (acquisition date) available. |
| **Post-conditions** | Cohort comparison chart shown. |
| **Typical Course of Action** | 1. Analyst requests cohort analysis. 2. System segments customers by acquisition date. 3. Retention/revenue trends displayed. |
| **Alternate Courses of Action** | Cohort by region. |
| **Exception Paths** | Insufficient cohort size. |

## UC-016 Inventory Risk Alerts

|  |  |
| --- | --- |
| **Identifier** | UC-016 |
| **Purpose** | Detect potential stockouts. |
| **Pre-conditions** | Inventory + sales velocity data available. |
| **Post-conditions** | Stockout alert triggered. |
| **Typical Course of Action** | 1. ERP Manager requests 'Flag stockout risks.' 2. System analyzes demand vs. inventory. 3. Alerts generated for SKUs. |
| **Alternate Courses of Action** | Alerts by warehouse. |
| **Exception Paths** | No stock data. |

## UC-017 Staffing vs. Sales Demand

|  |  |
| --- | --- |
| **Identifier** | UC-017 |
| **Purpose** | Recommend staffing adjustments. |
| **Pre-conditions** | Staffing + sales forecast data available. |
| **Post-conditions** | Staffing adjustment report shown. |
| **Typical Course of Action** | 1. ERP Manager asks: 'Do I need more staff next month?' 2. System forecasts workload. 3. Staffing gap recommendations displayed. |
| **Alternate Courses of Action** | Recommendations by department. |
| **Exception Paths** | Staffing data missing. |

## UC-018 Supply Chain Stress Test

|  |  |
| --- | --- |
| **Identifier** | UC-018 |
| **Purpose** | Simulate impact of demand surge. |
| **Pre-conditions** | Supply chain data linked. |
| **Post-conditions** | Stress test results shown. |
| **Typical Course of Action** | 1. ERP Manager requests: 'Simulate 15% demand surge in East.' 2. System models logistics and capacity. 3. Stress report displayed. |
| **Alternate Courses of Action** | Change surge % value. |
| **Exception Paths** | Missing logistics data. |

## UC-019 Lead Source ROI Comparison

|  |  |
| --- | --- |
| **Identifier** | UC-019 |
| **Purpose** | Compare ROI of lead generation channels. |
| **Pre-conditions** | Lead source and conversion data available. |
| **Post-conditions** | ROI table displayed. |
| **Typical Course of Action** | 1. Marketing Strategist asks: 'Compare LinkedIn vs. Email ROI.' 2. System links lead source to revenue outcomes. 3. ROI table and chart displayed. |
| **Alternate Courses of Action** | Compare across multiple sources. |
| **Exception Paths** | Incomplete lead source attribution. |

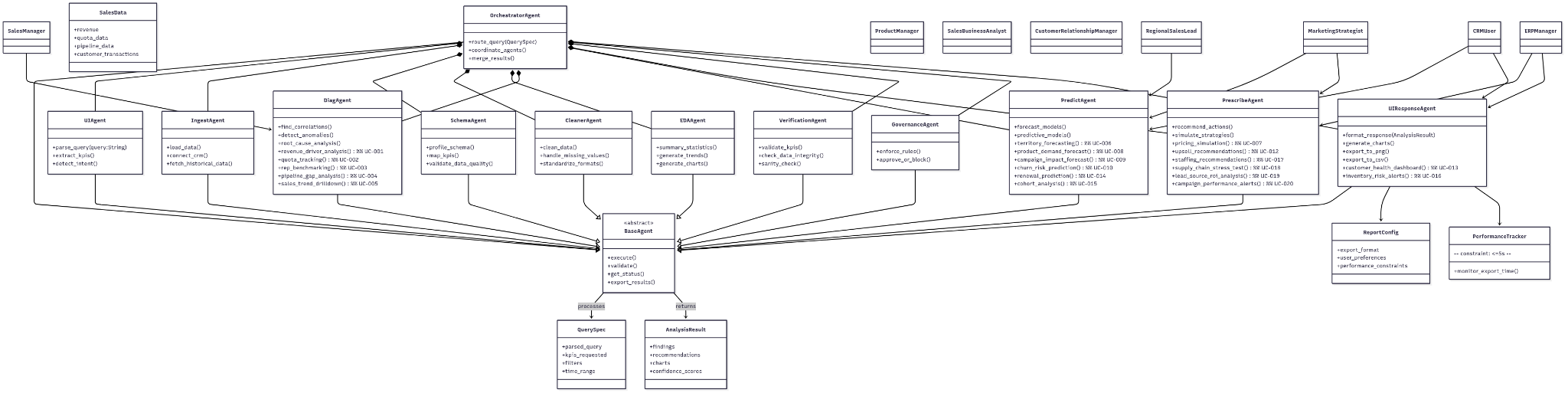
## UC-020 Campaign Underperformance Alerts

|  |  |
| --- | --- |
| **Identifier** | UC-020 |
| **Purpose** | Alert when campaign ROI falls below benchmark. |
| **Pre-conditions** | Campaign metrics available. |
| **Post-conditions** | Alert triggered. |
| **Typical Course of Action** | 1. Strategist sets performance thresholds. 2. System monitors campaign ROI. 3. If ROI < threshold, alert generated. |
| **Alternate Courses of Action** | Weekly summary instead of real-time alerts. |
| **Exception Paths** | No active campaigns. |

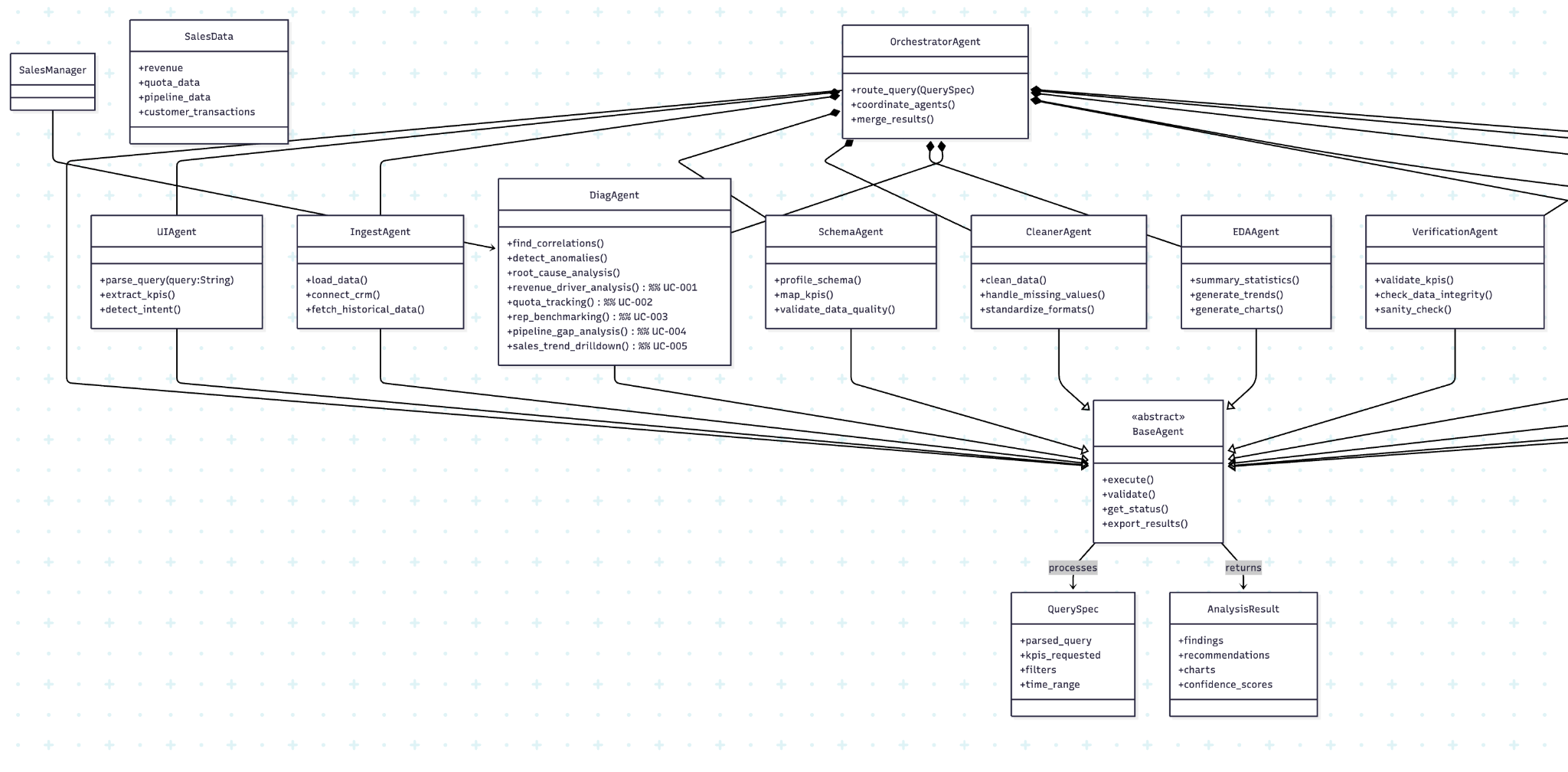
# Class Diagram

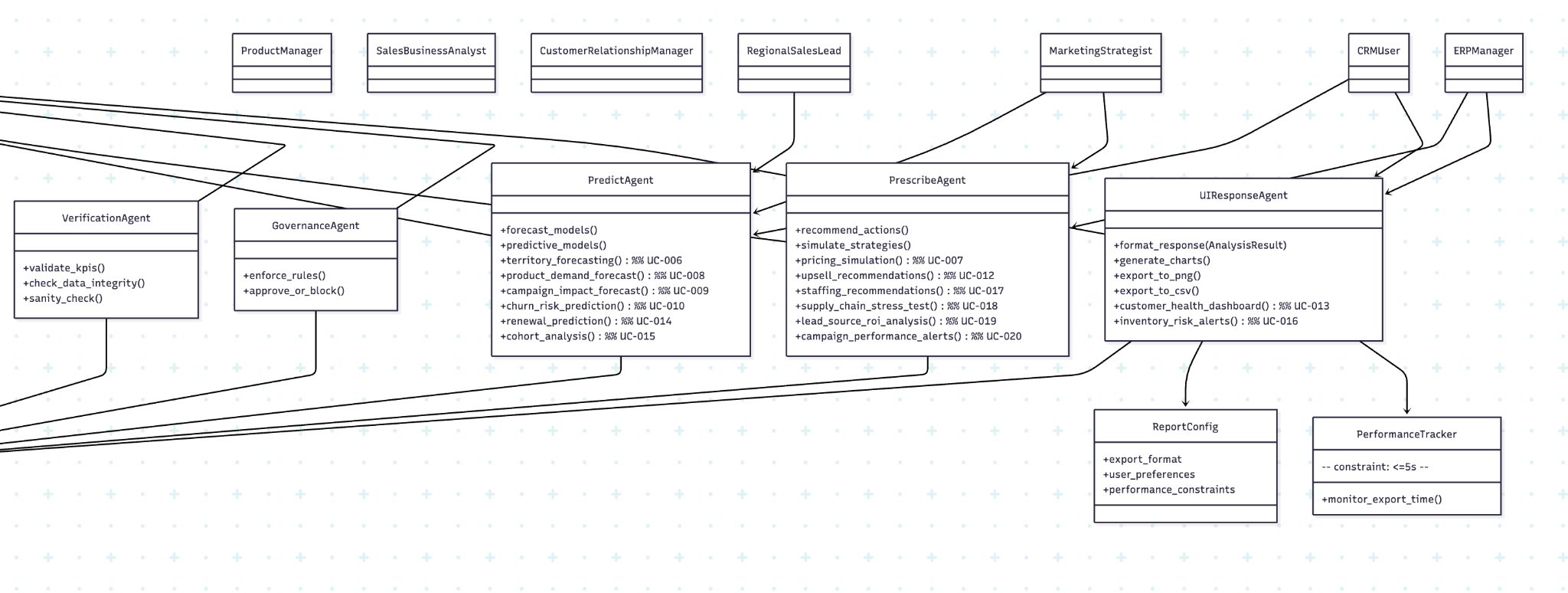
Add relation between classes like inheritance, aggregation etc.

## Diagram



Zoomed in:





## Description

## Abstract Layer

### **BaseAgent**

**Purpose**: Abstract parent class for all agents. Provides standard lifecycle operations.  
 **Operations**:

* execute() – run assigned task
* validate() – ensure inputs and outputs are correct
* getStatus() – check agent health
* exportResults() – return results in a standard format

## Core Agents

### **UIAgent**

**Purpose**: Interprets user queries in natural language. Extracts KPIs, filters, and intent.  
 **Operations**:

* parseQuery(query: String)
* extractKPIs()
* detectIntent()

### **OrchestratorAgent**

**Purpose**: Central controller that routes queries, invokes agents in the right order, and merges results.  
 **Operations**:

* routeQuery
* coordinateAgents()
* mergeResults()

### **IngestAgent**

**Purpose**: Connects to CRM/ERP/databases and loads raw data.  
 **Operations**:

* loadData()
* connectCRM()
* fetchHistoricalData()

### **SchemaAgent**

**Purpose**: Profiles datasets, maps columns to KPIs, and validates schema quality.  
 **Operations**:

* profileSchema()
* mapKPIs()
* validateDataQuality()

### **CleanerAgent**

**Purpose**: Cleans and standardizes raw data. Handles missing values and format inconsistencies.  
 **Operations**:

* cleanData()
* handleMissingValues()
* standardizeFormats()

### **EDAAgent**

**Purpose**: Performs exploratory data analysis, including descriptive statistics and trends.  
 **Operations**:

* summaryStatistics()
* generateTrends()
* generateCharts()

### **DiagAgent**

**Purpose**: Diagnostic analysis to identify root causes, anomalies, and correlations.  
 **Operations (Use Cases)**:

* revenueDriverAnalysis() – UC-001
* quotaTracking() – UC-002
* repBenchmarking() – UC-003
* pipelineGapAnalysis() – UC-004
* salesTrendDrilldown() – UC-005

### **PredictAgent**

**Purpose**: Predictive analytics and forecasting using models.  
 **Operations (Use Cases)**:

* territoryForecasting() – UC-006
* productDemandForecast() – UC-008
* campaignImpactForecast() – UC-009
* churnRiskPrediction() – UC-010
* renewalPrediction() – UC-014
* cohortAnalysis() – UC-015

### **PrescribeAgent**

**Purpose**: Prescriptive analytics that generates actionable recommendations and simulations.  
 **Operations (Use Cases)**:

* pricingSimulation() – UC-007
* upsellRecommendations() – UC-012
* staffingRecommendations() – UC-017
* supplyChainStressTest() – UC-018
* leadSourceROIAnalysis() – UC-019
* campaignPerformanceAlerts() – UC-020

### **VerificationAgent**

**Purpose**: Validates outputs to ensure KPIs and results are correct.  
 **Operations**:

* validateKPIs()
* checkDataIntegrity()
* sanityCheck()

### **GovernanceAgent**

**Purpose**: Enforces business and compliance rules before results are shown.  
 **Operations**:

* enforceRules()
* approveOrBlock()

### **UIResponseAgent**

**Purpose**: Formats and delivers results to the user, including visualizations and exports.  
 **Operations (Use Cases)**:

* formatResponse(result: AnalysisResult)
* generateCharts()
* exportToPNG()
* exportToCSV()
* customerHealthDashboard() – UC-013
* inventoryRiskAlerts() – UC-016

## Data Model Classes

### **SalesData**

**Attributes**:

* revenue
* quotaData
* pipelineData
* customerTransactions  
   **Purpose**: Stores sales datasets for analysis.

### **QuerySpec**

**Attributes**:

* parsedQuery
* kpisRequested
* filters
* timeRange  
   **Purpose**: Structured query specification passed from UIAgent to Orchestrator.

### **AnalysisResult**

**Attributes**:

* findings
* recommendations
* charts
* confidenceScores  
   **Purpose**: Standardized result object returned by agents.

### **ReportConfig**

**Attributes**:

* exportFormat
* userPreferences
* performanceConstraints  
   **Purpose**: Controls how results are exported.

### **PerformanceTracker**

**Operations**:

* monitorExportTime()  
   **Constraint**: All exports must complete within 5 seconds.

## User Role Classes

### **SalesManager**

**Purpose**: Uses sales diagnostic and trend analysis (UC-001 → UC-005).

### **RegionalLead**

**Purpose**: Focuses on territory forecasting (UC-006).

### **ERPManager**

**Purpose**: Manages forecasting, inventory, staffing, and supply chain use cases (UC-008, UC-016, UC-017, UC-018).

### **CRMUser**

**Purpose**: Works on customer management and retention (UC-010 → UC-015).

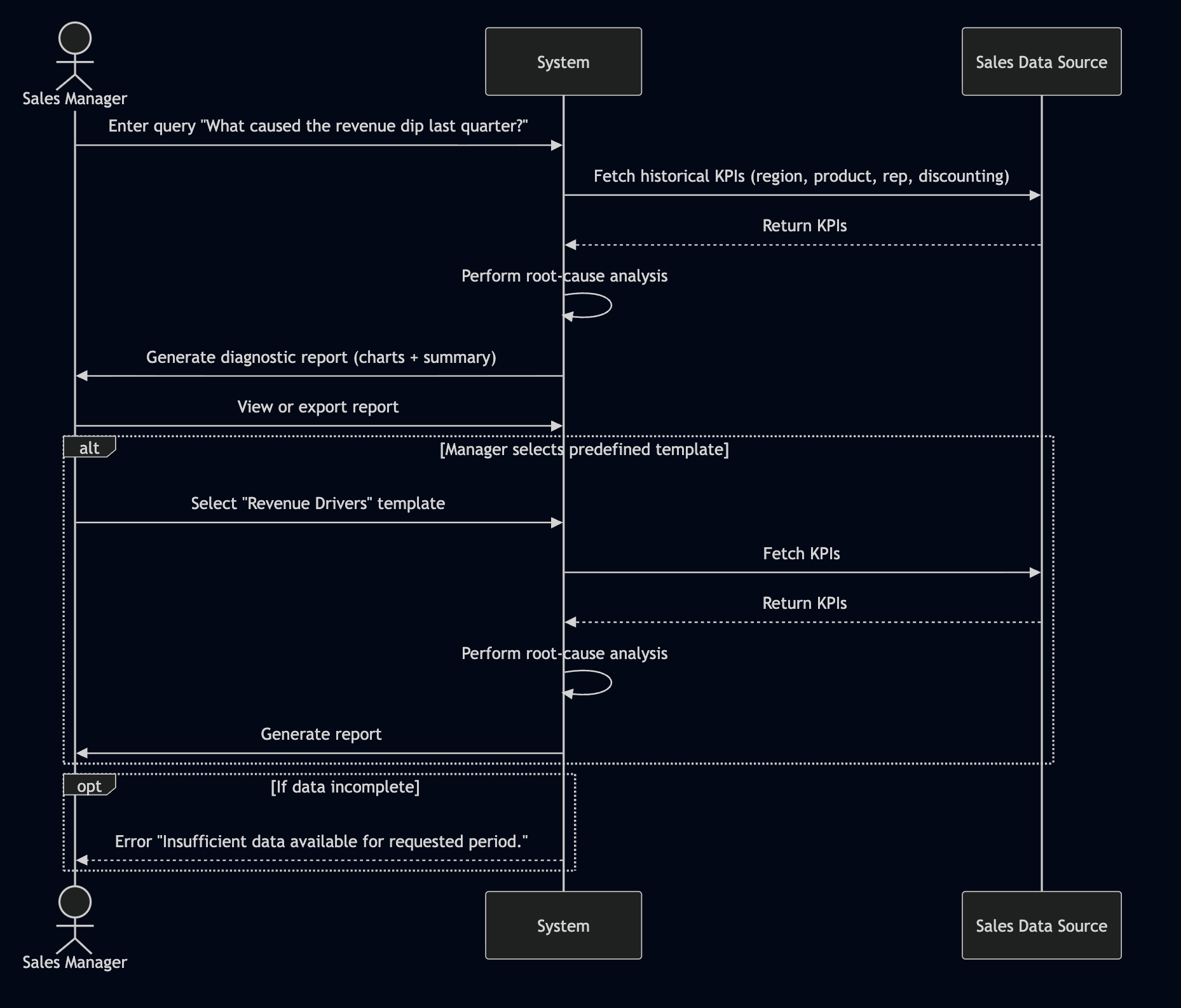
### **MarketingStrategist**

**Purpose**: Manages campaign impact and ROI analysis (UC-009, UC-019, UC-020).

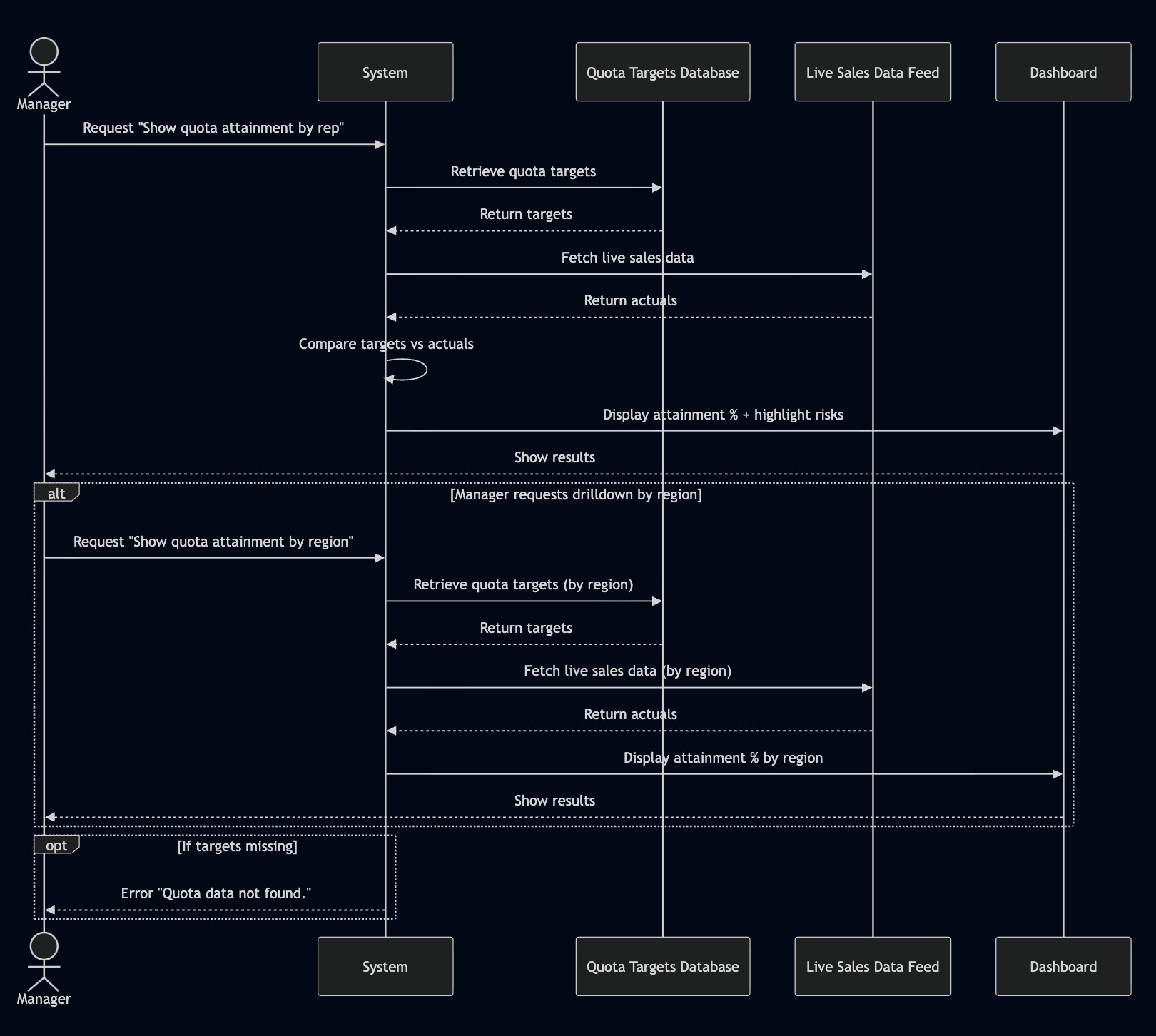
# Sequence Diagrams

Update this section, use the above classes in the sequence diagrams.

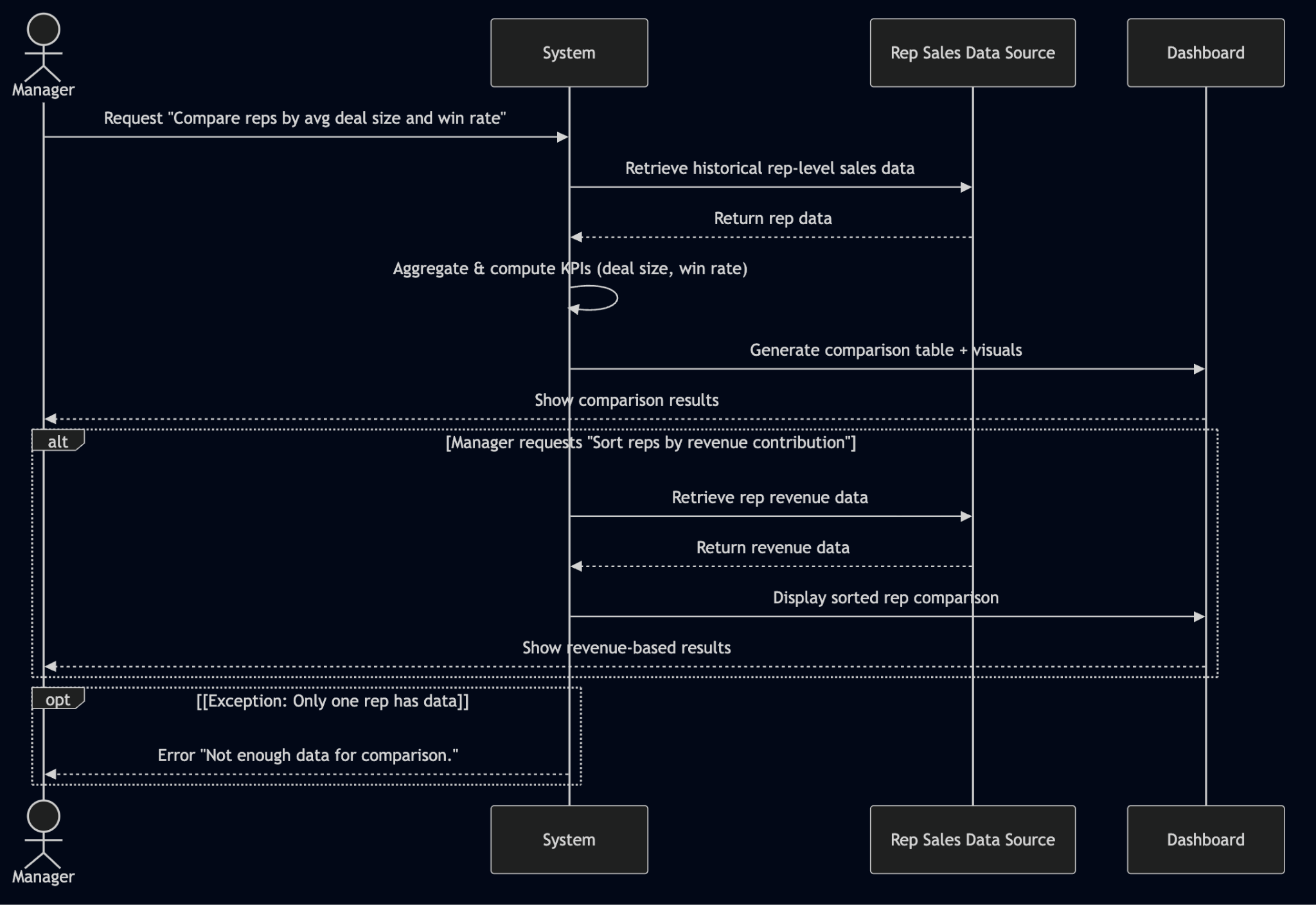
## UC-001 Revenue Driver Analysis



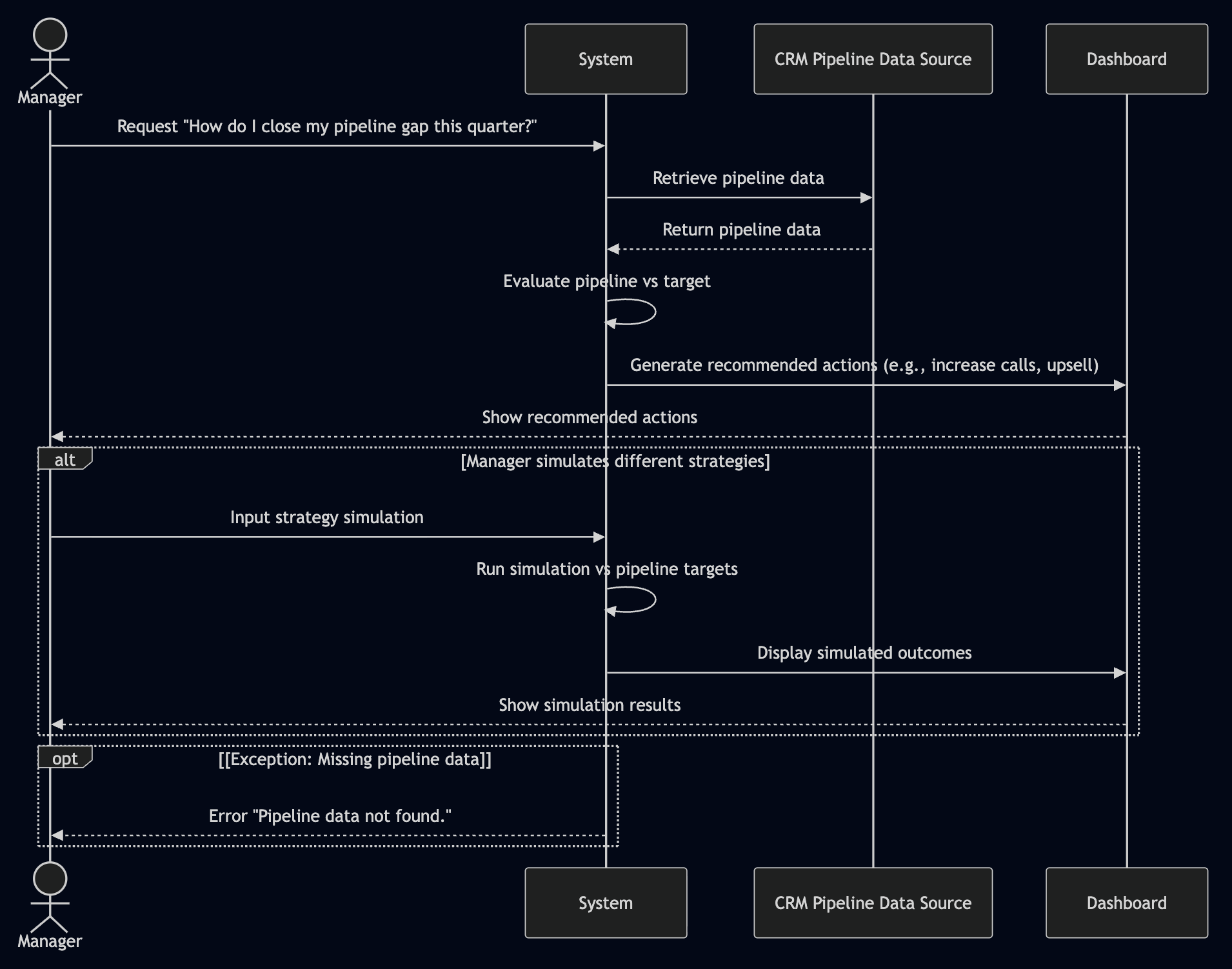
## UC - 002 Quota Tracking



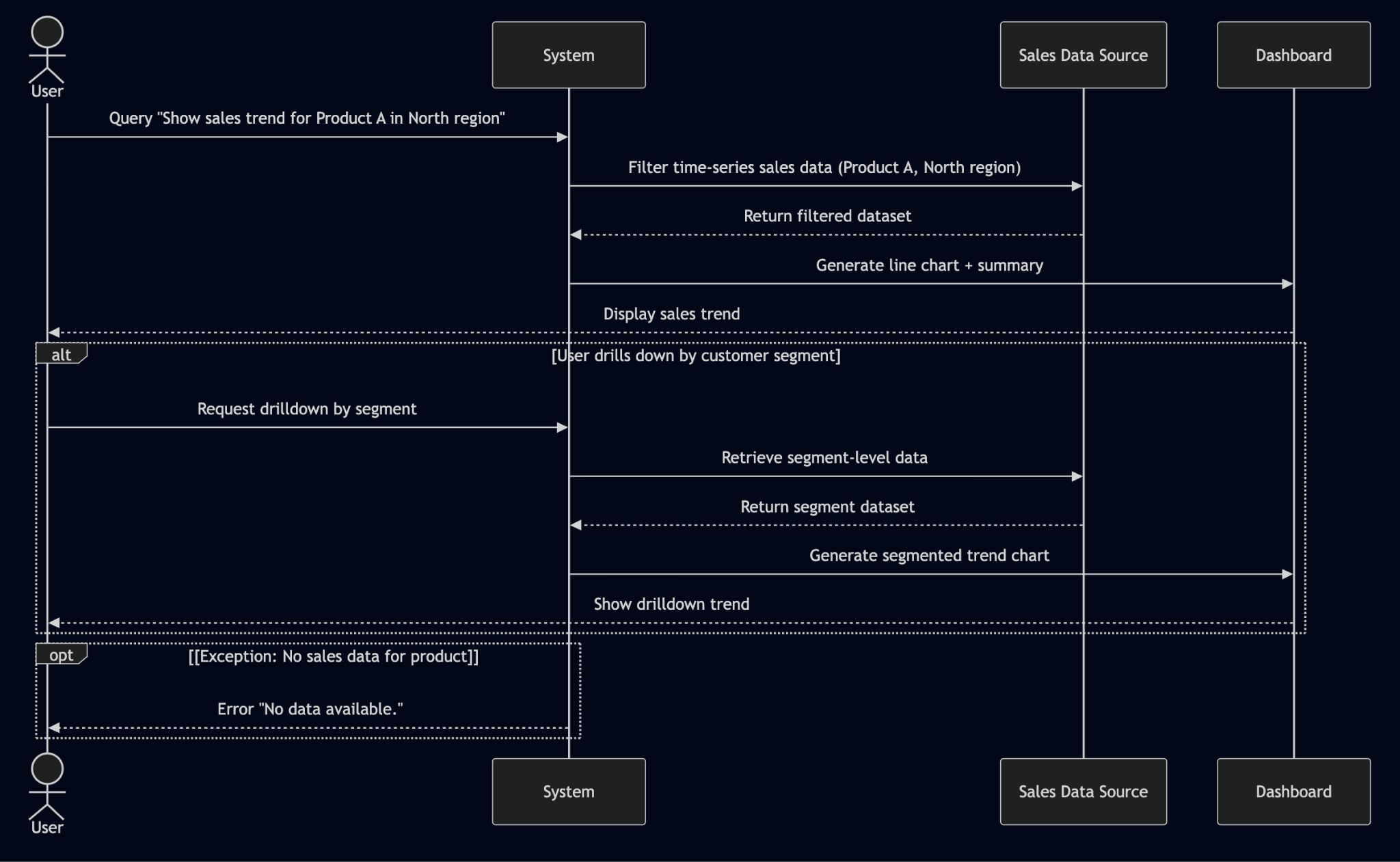
## UC - 003 Rep Bench Marking



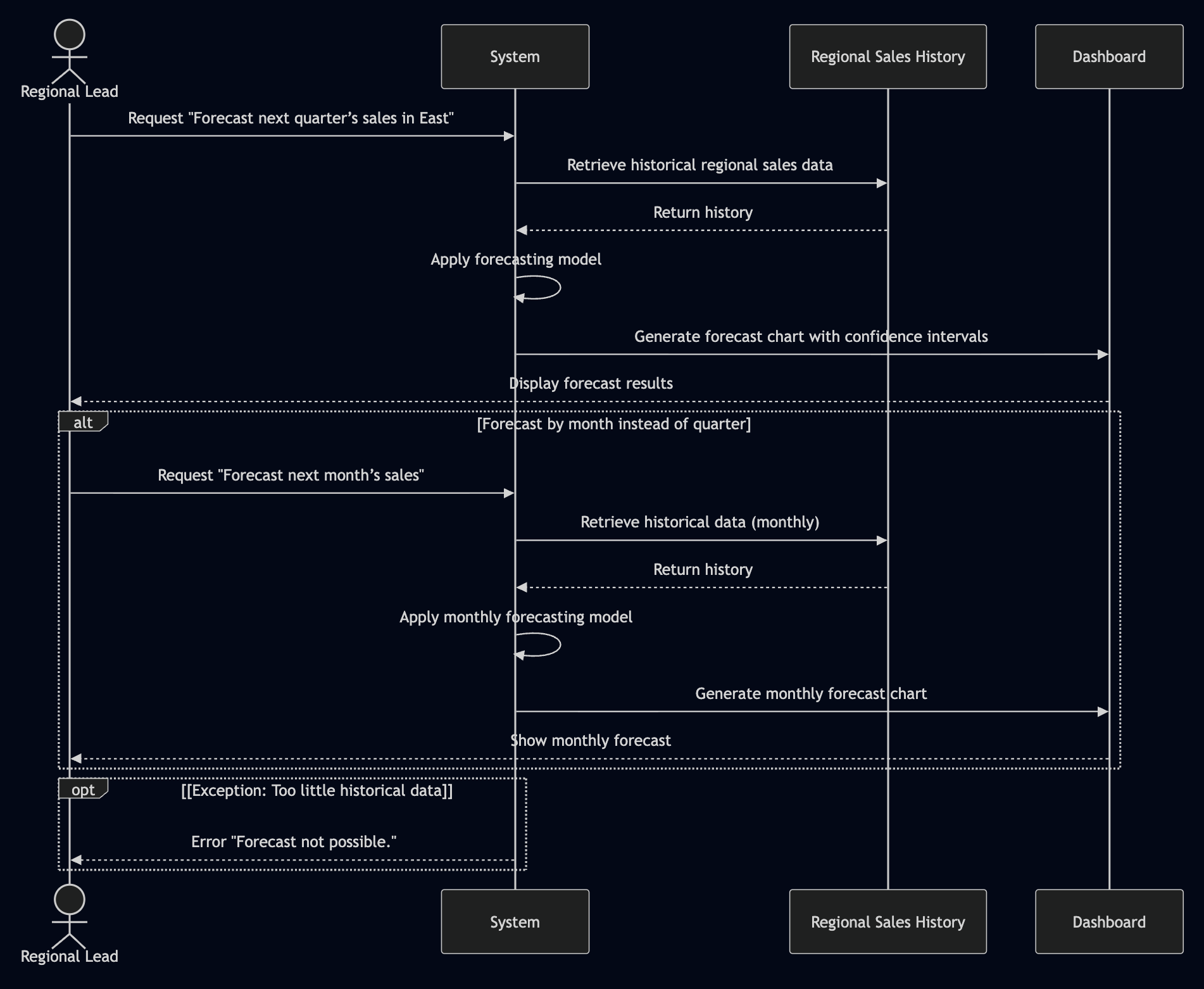
## UC - 004 Pipeline Gap Closure



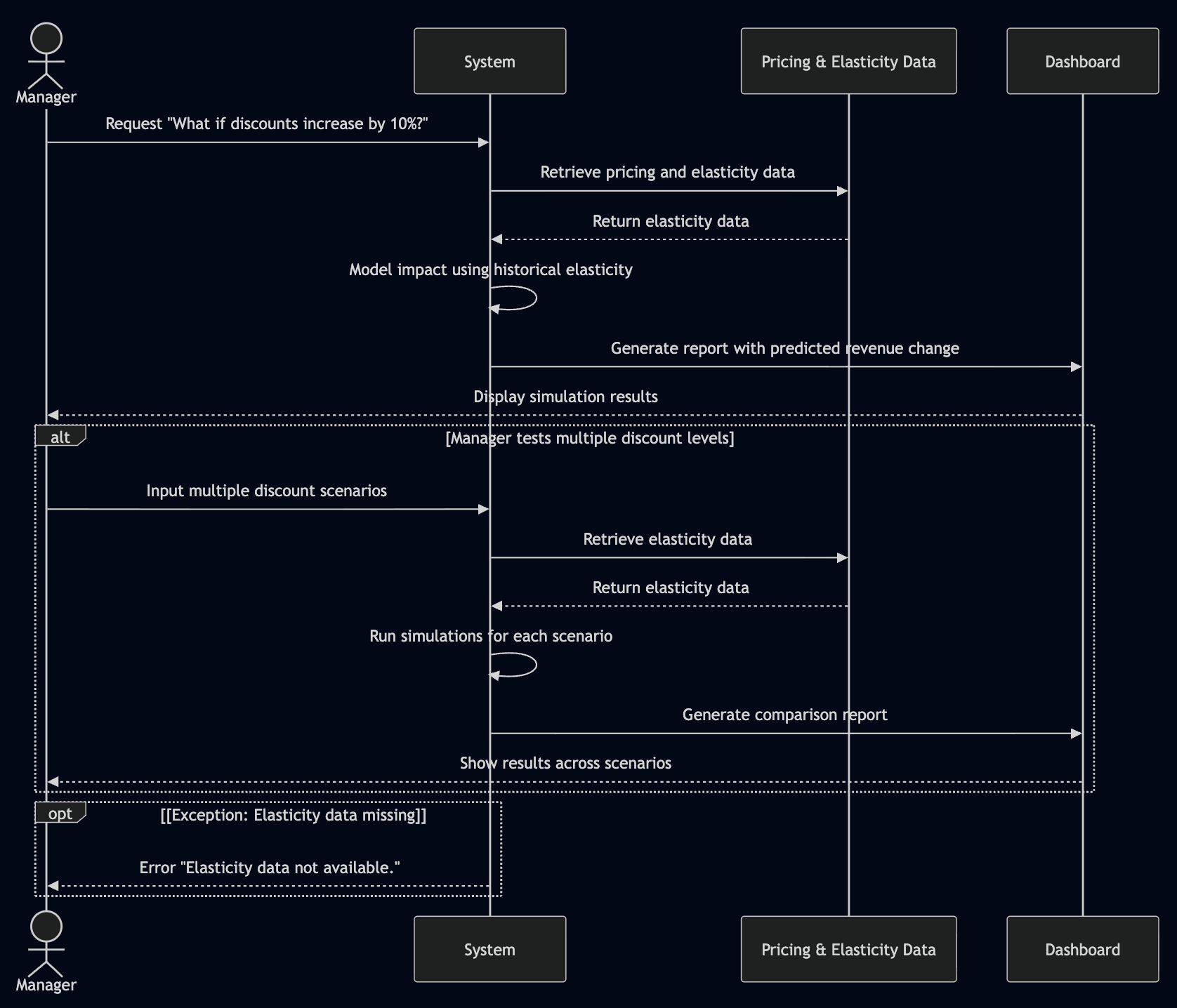
## UC - 005 Sales Trend Drilldown



## UC - 006 Territory Forecasting



## UC - 007 What If Price Simulation



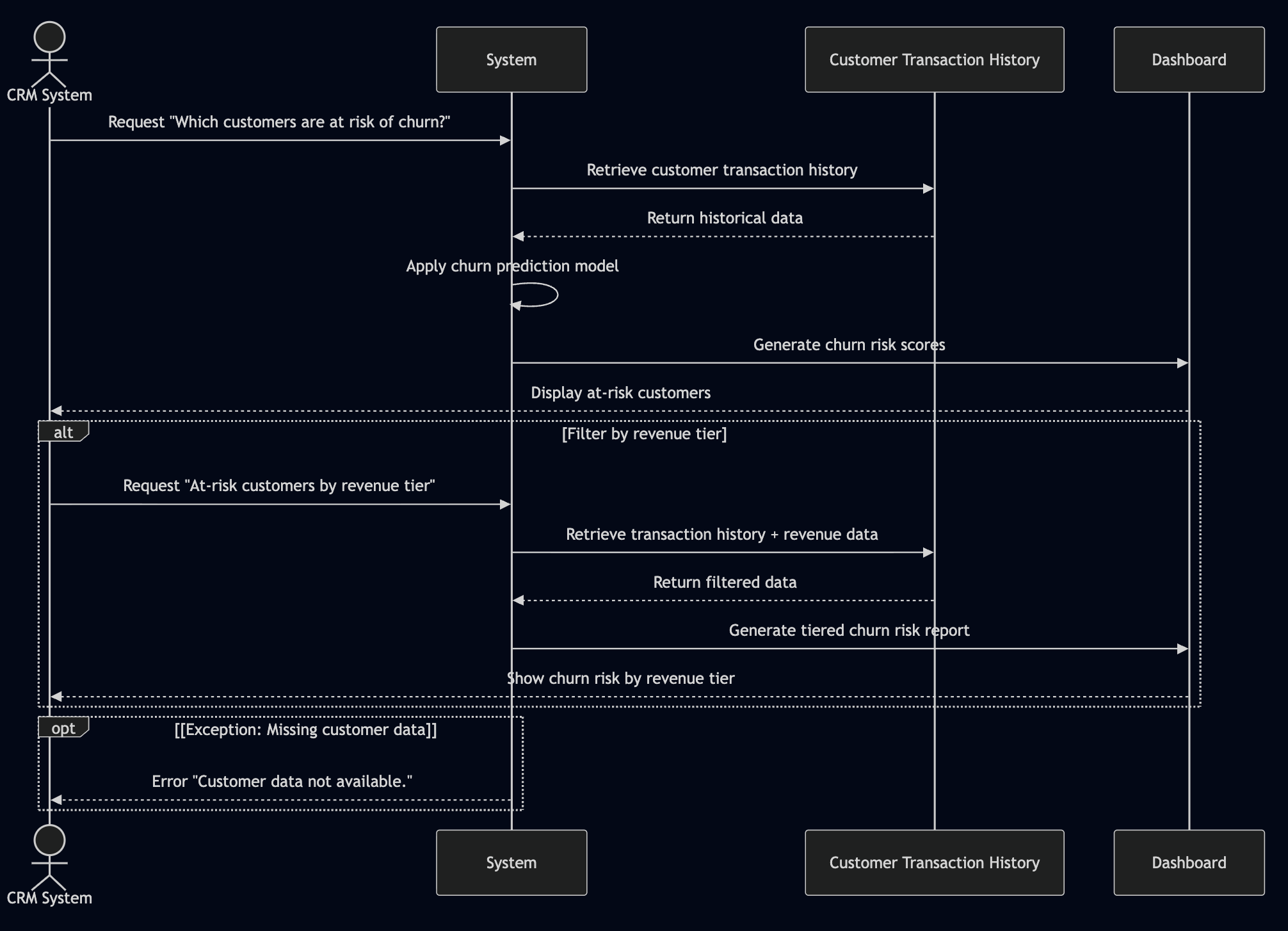
## UC - 008 Product Demand Forecast



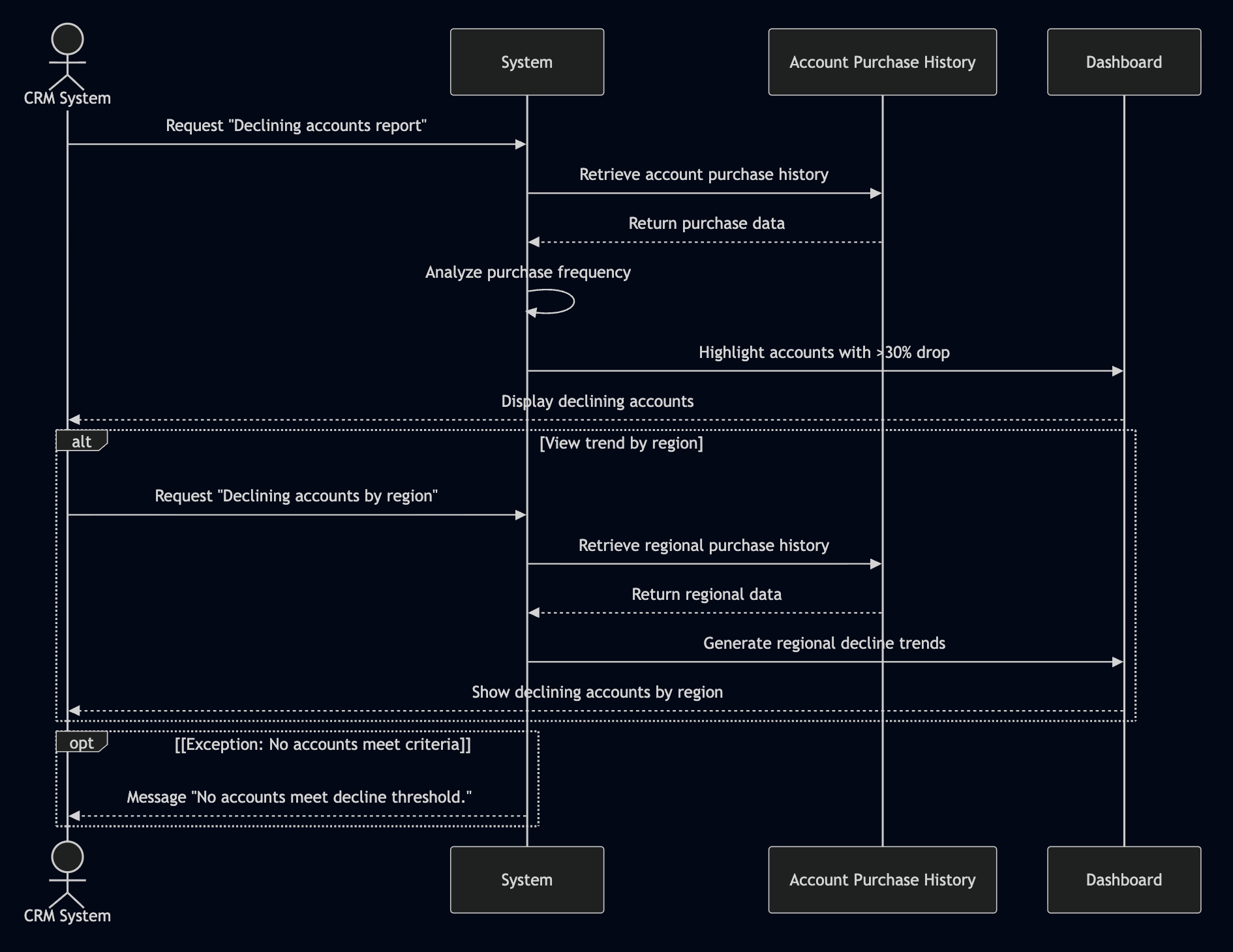
## UC - 009 Campaign Impact Forecast



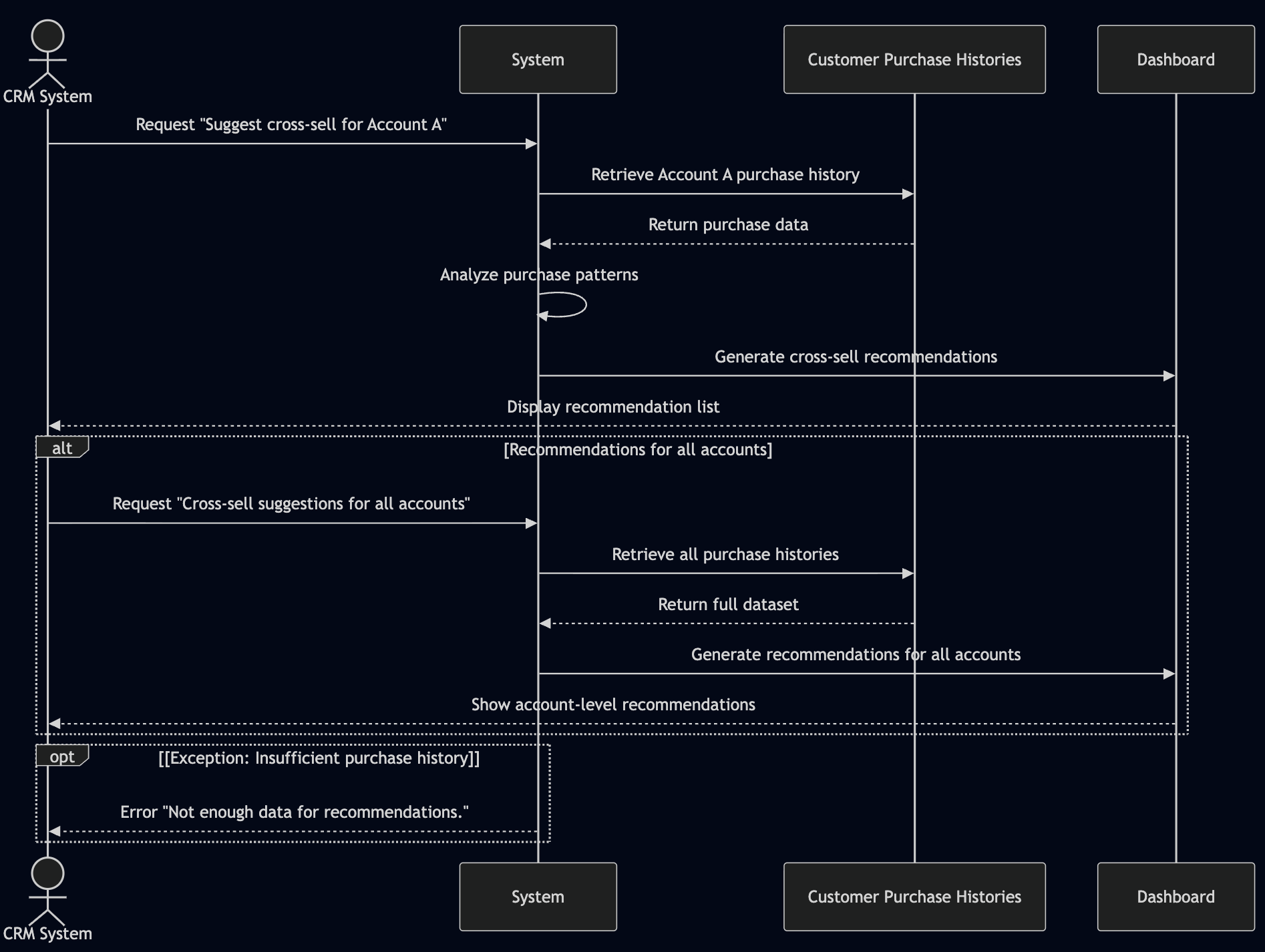
## UC - 010 Churn Risk Forecasting



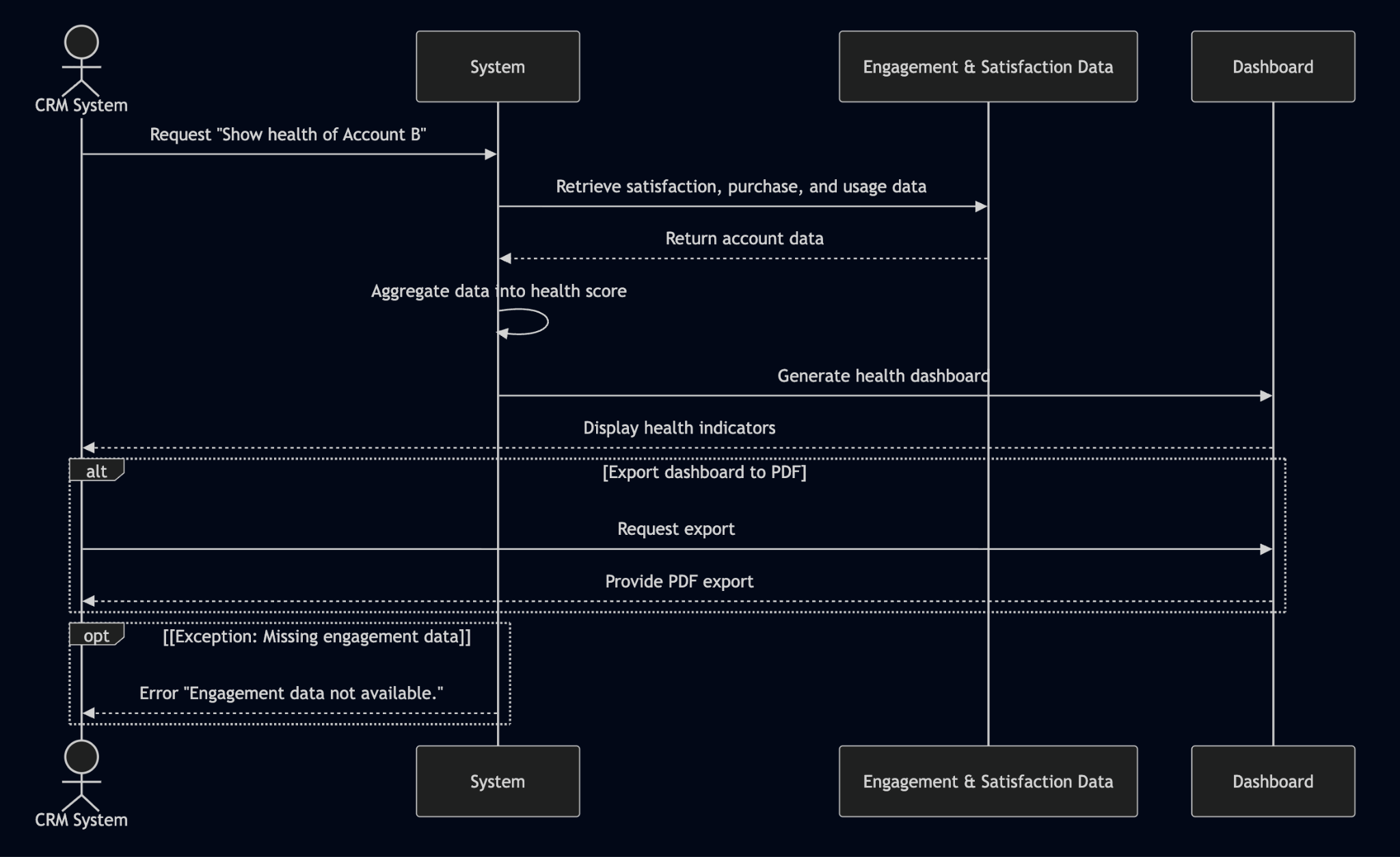
## UC - 011 At-Risk Accounts



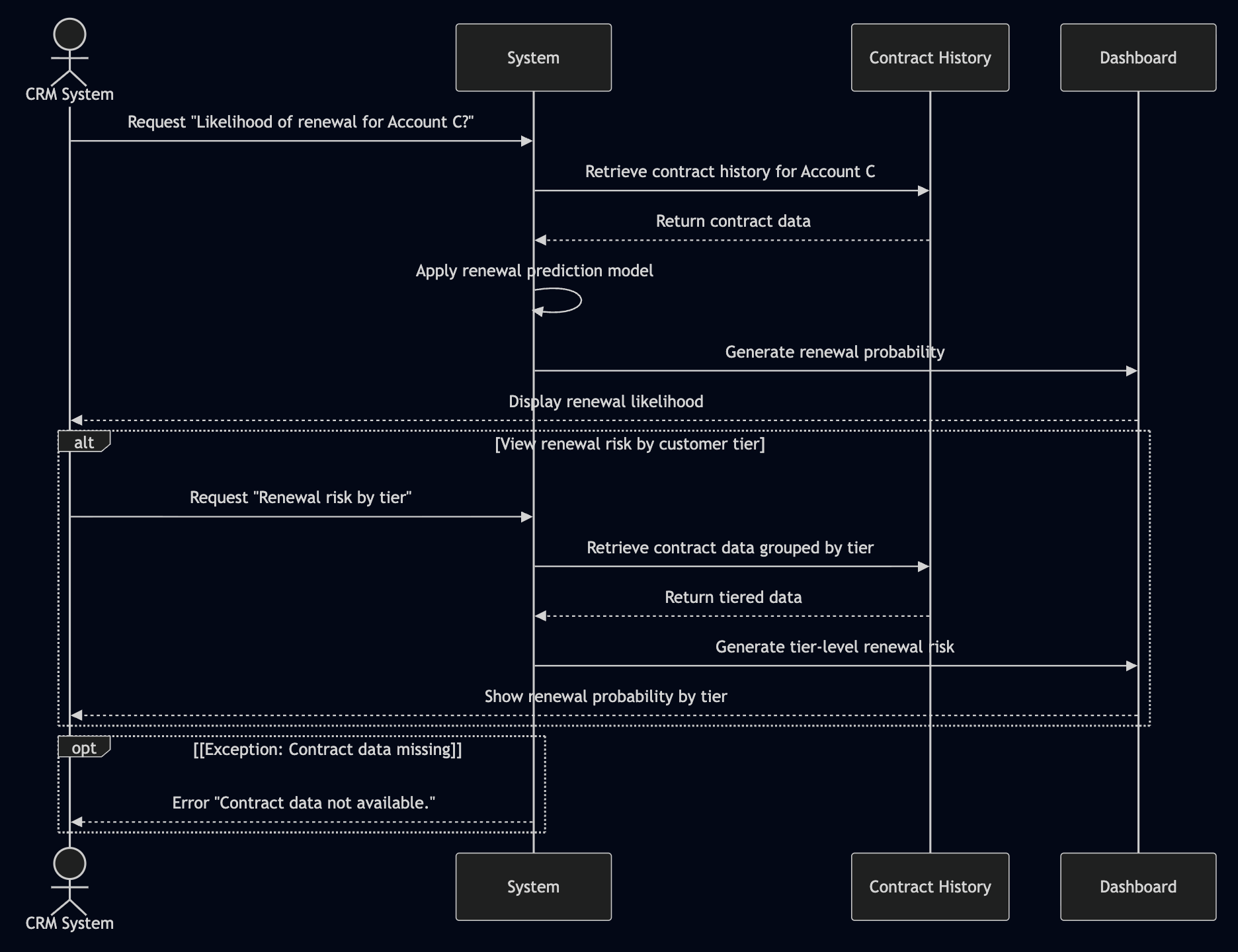
## UC - 012 Upsell/Cross-Sell Suggestions



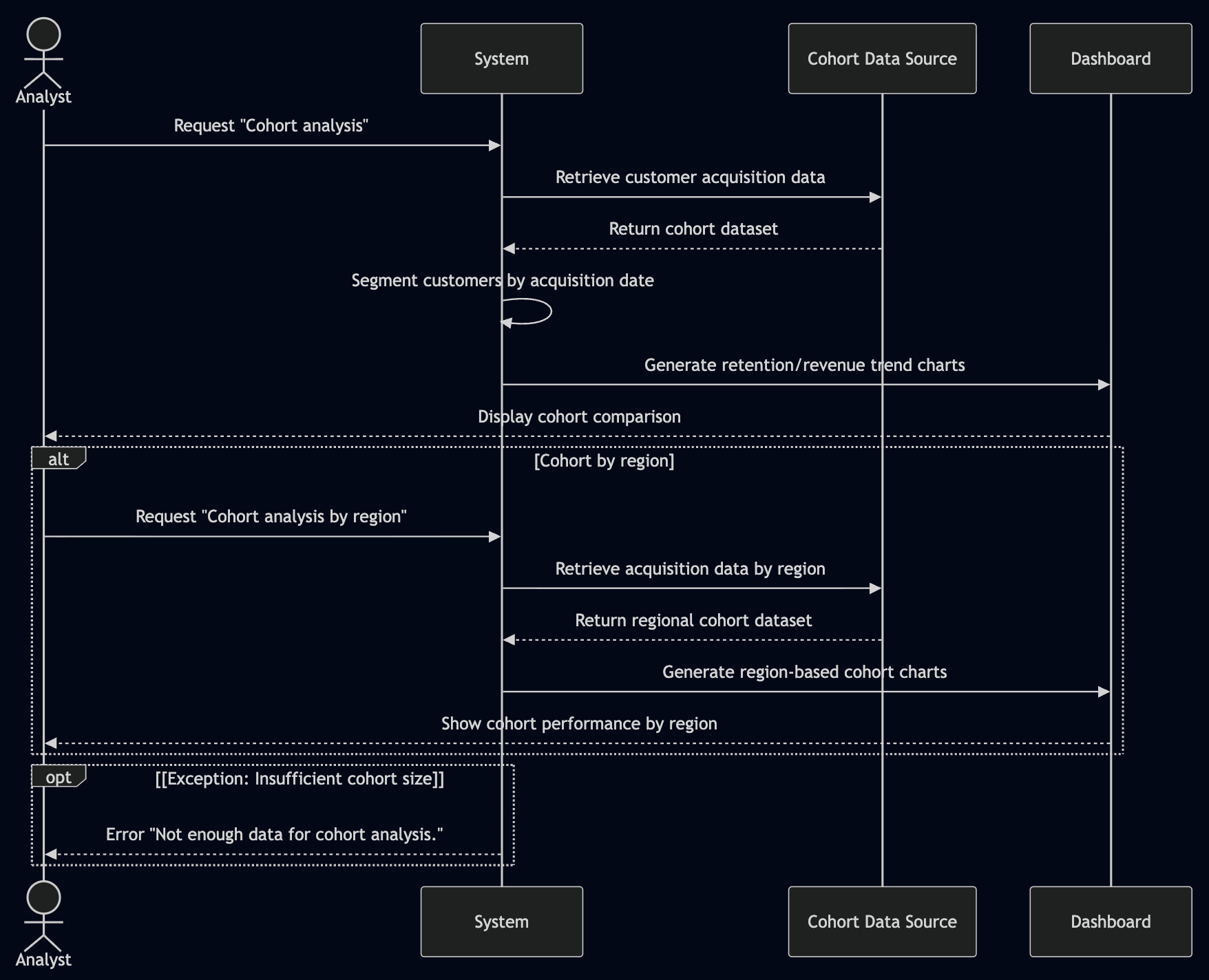
## UC - 013 Customer Health Dashboard



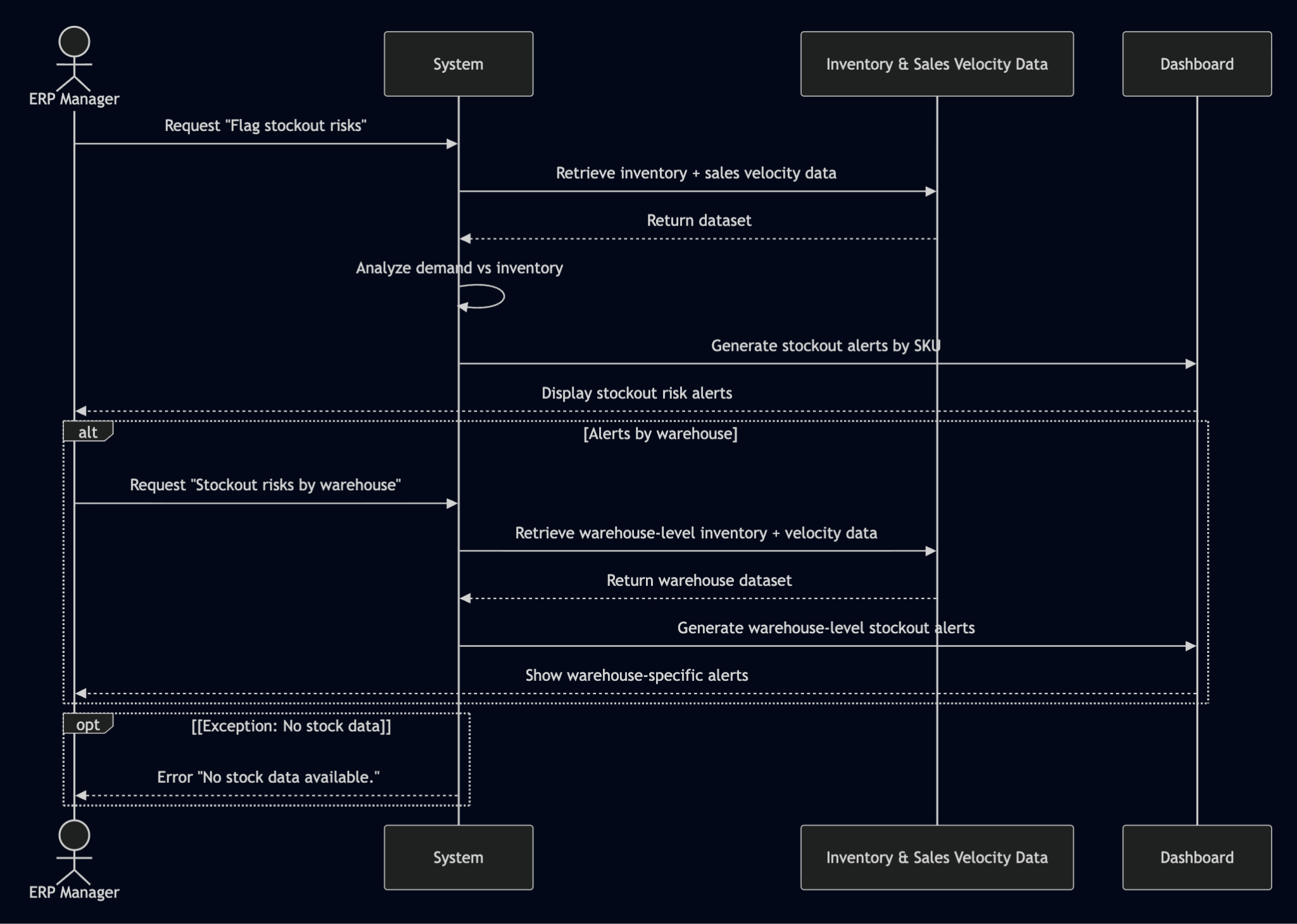
## UC - 014 Renewal Likelihood Prediction



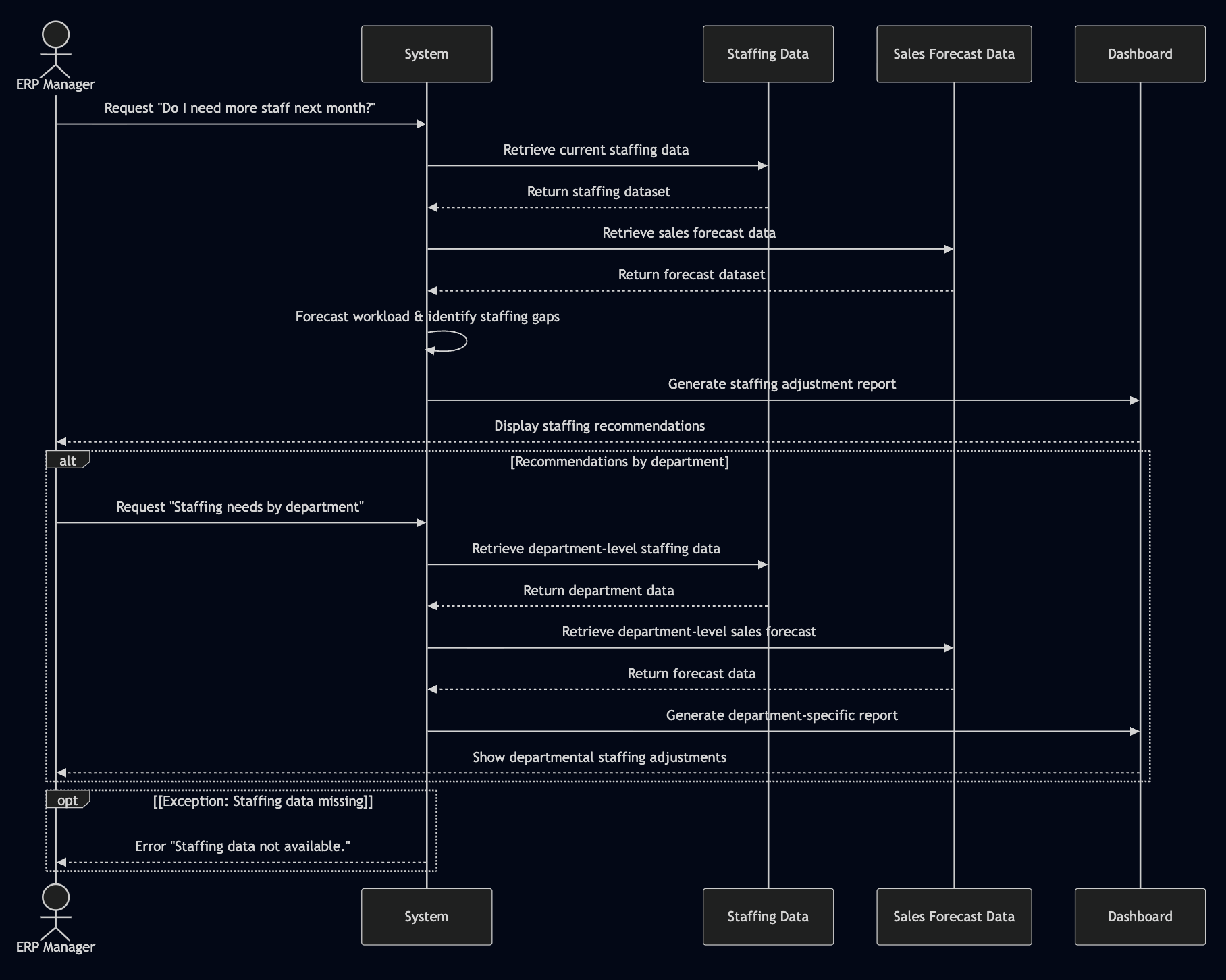
## UC - 015 Customer Cohort Analysis



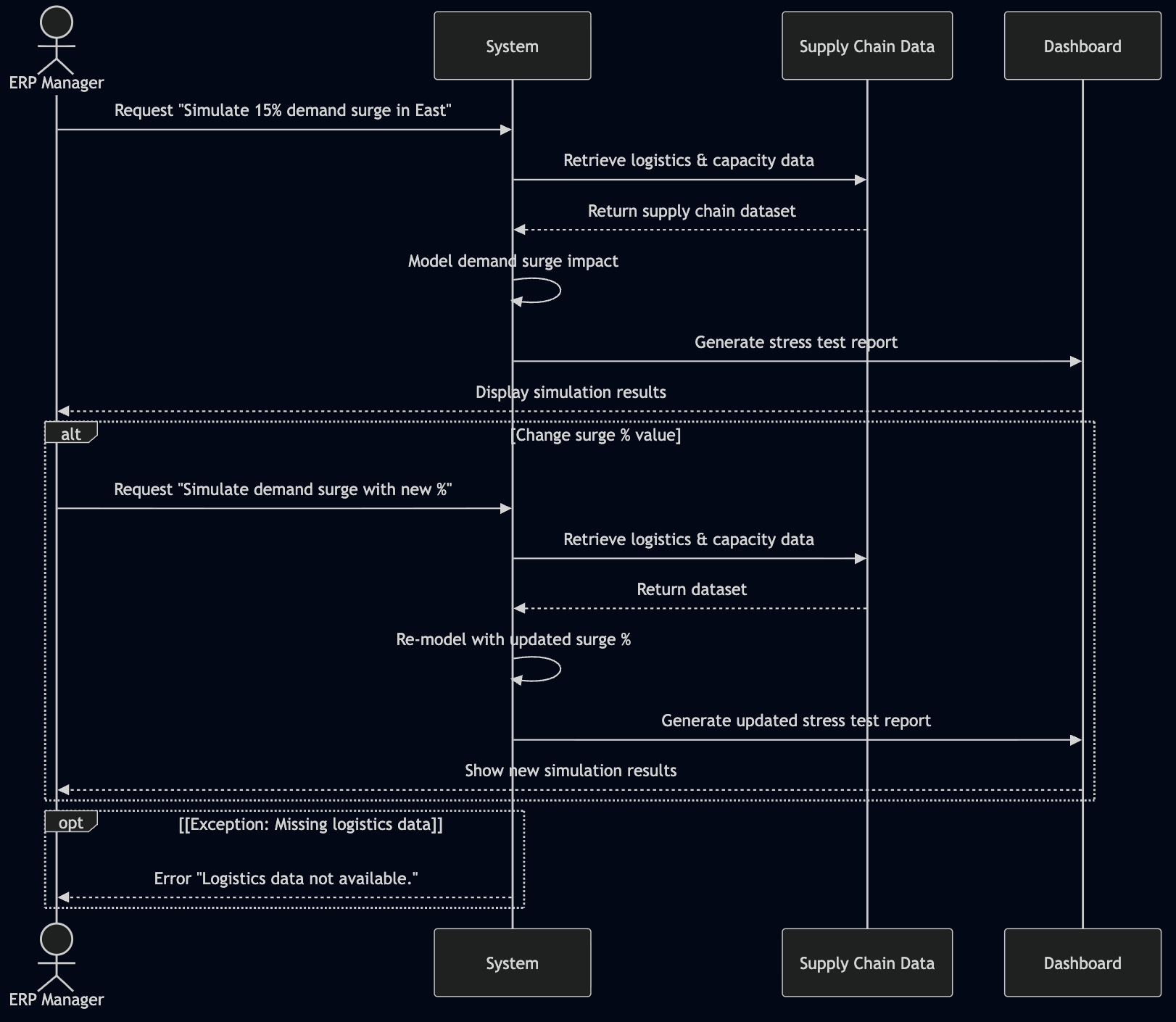
## UC - 016 Inventory Risk Alerts



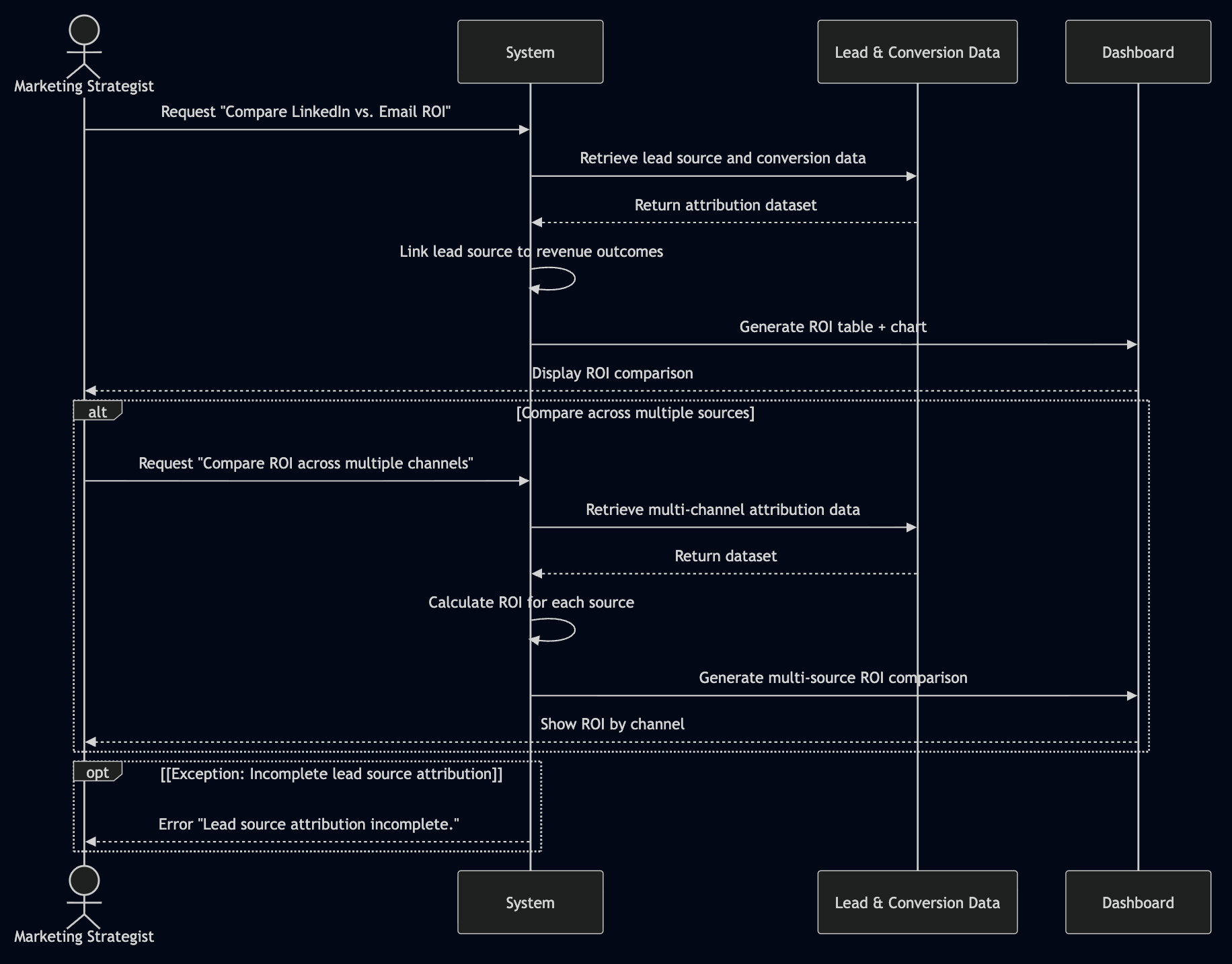
## UC - 017 Staffing vs Sales Demand



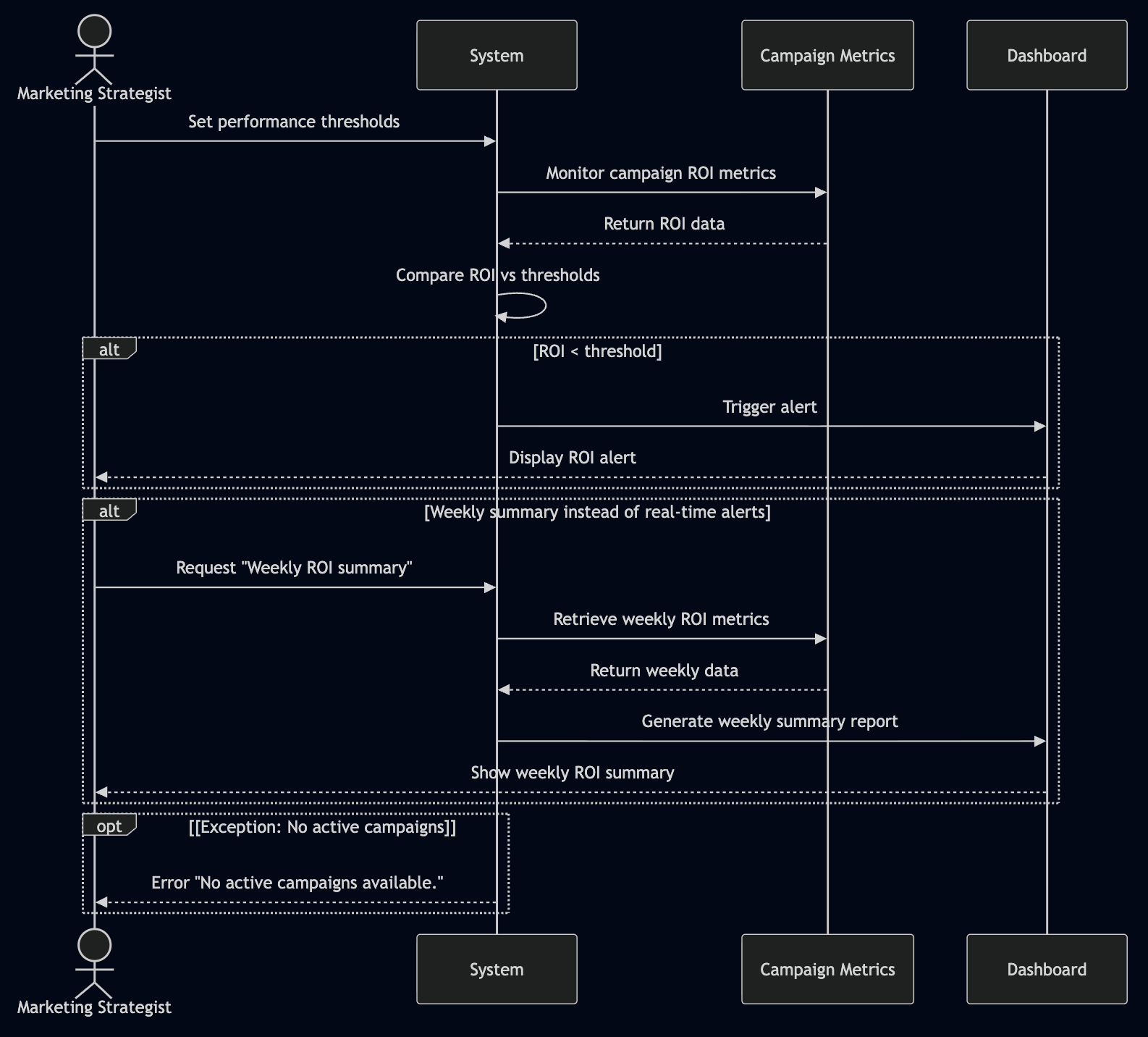
## UC - 018 Supply Chain Stress Test



## UC - 019 Lead Source ROI Comparison



## UC - 020 Campaign Underperformance Alerts



# State Diagrams

## Diagram details

## Object: Query Processing Engine

**Details:** Core engine that processes natural language business queries through multiple analytical stages.

### **States and Transitions**

* **Waiting:** Ready to receive new query
* **Parsing Query:** Understanding natural language input and extracting intent
* **Validating Requirements:** Checking if query can be processed with available data
* **Requesting Clarification:** Asking user for additional information when query is ambiguous
* **Data Retrieval:** Fetching relevant datasets based on query requirements
* **Analysis Processing:** Applying appropriate analytical methods (descriptive, predictive, prescriptive)
* **Report Generation:** Creating visualizations, charts, and summary reports
* **Results Ready:** Final results prepared for user presentation
* **Error State:** Handling exceptions like missing data, insufficient history, etc.

### **Key Messages/Events**

* **receiveQuery():** transitions to Parsing Query
* **queryParsed():** transitions to Validating Requirements
* **requirementsMet():** transitions to Data Retrieval
* **needsClarification():** transitions to Requesting Clarification
* **clarificationReceived():** transitions to Data Retrieval
* **dataRetrieved():** transitions to Analysis Processing
* **analysisComplete():** transitions to Report Generation
* **reportGenerated():** transitions to Results Ready
* **error():** transitions to Error State
* **reset():** transitions back to Waiting

## Object: Data Validator

**Details:** Validates data availability and quality before processing user queries; handles various error conditions.

### **States and Transitions**

* **Ready:** Waiting for data validation request
* **Checking Availability:** Verifying if required data exists for the query
* **Checking Completeness:** Ensuring data is sufficient for analysis
* **Validated:** Data passes all checks
* **Insufficient Data:** Not enough data for requested analysis
* **Missing Dataset:** Required dataset not found
* **Data Quality Issues:** Data exists but has quality problems
* **Recovery Mode:** Attempting to find alternative data sources

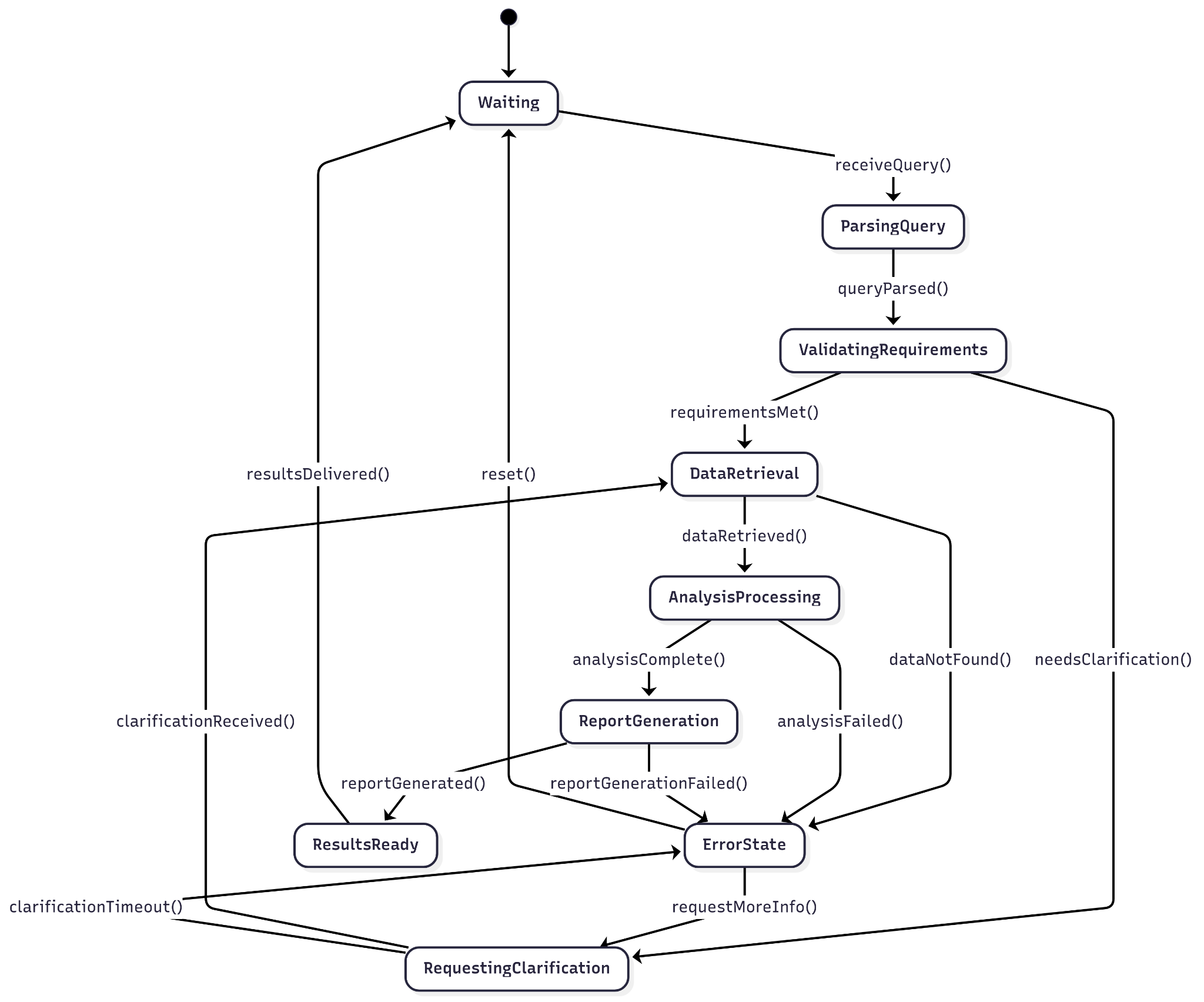
### **Key Messages/Events**

* **validateData():** transitions to Checking Availability
* **dataExists():** transitions to Checking Completeness
* **dataComplete():** transitions to Validated
* **insufficientData():** transitions to Insufficient Data
* **datasetMissing():** transitions to Missing Dataset
* **qualityIssues():** transitions to Data Quality Issues
* **attemptRecovery():** transitions to Recovery Mode

## Diagram

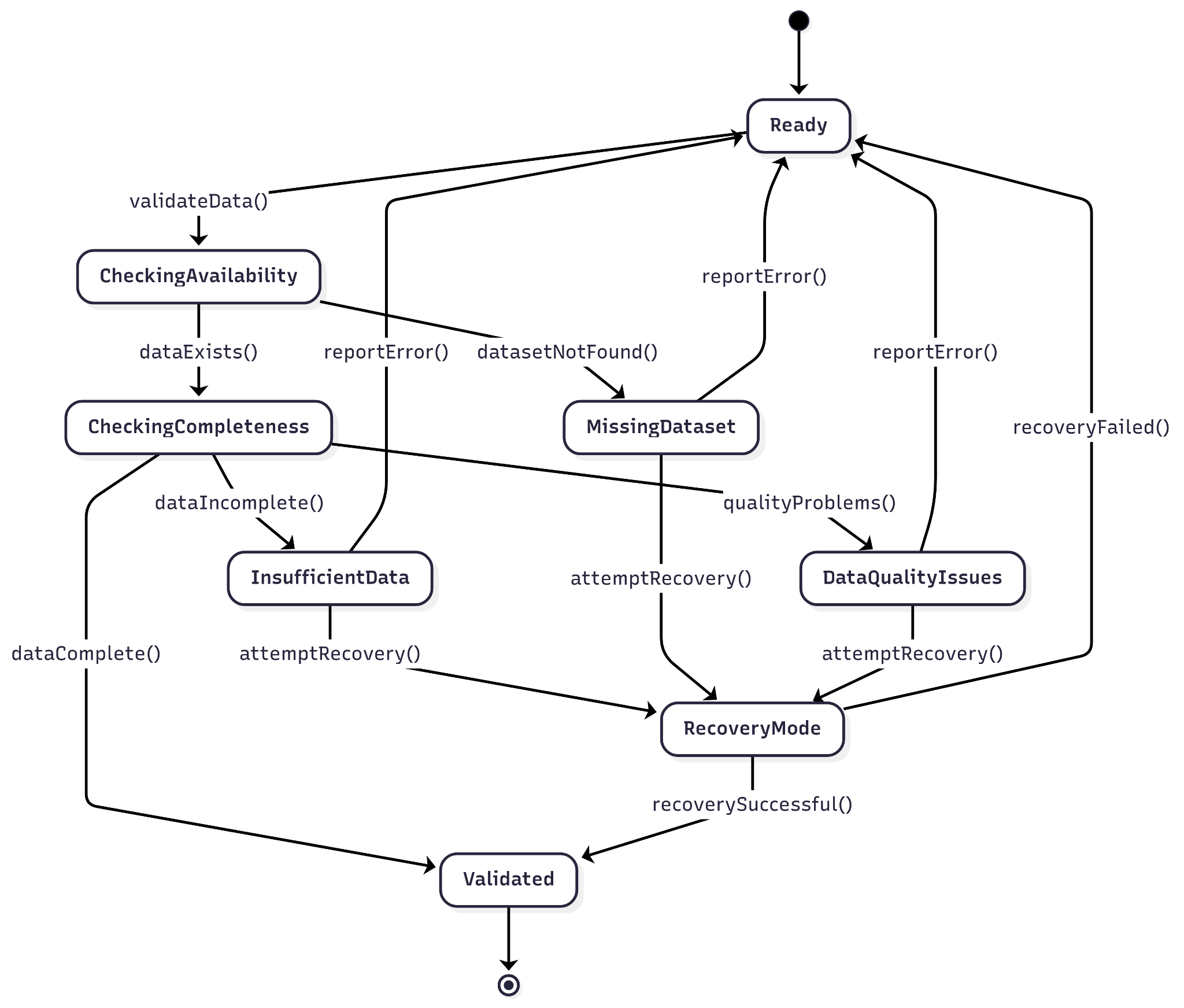
Add final state according to uml notation.

## Query Processing Engine



…..

## Data Validator



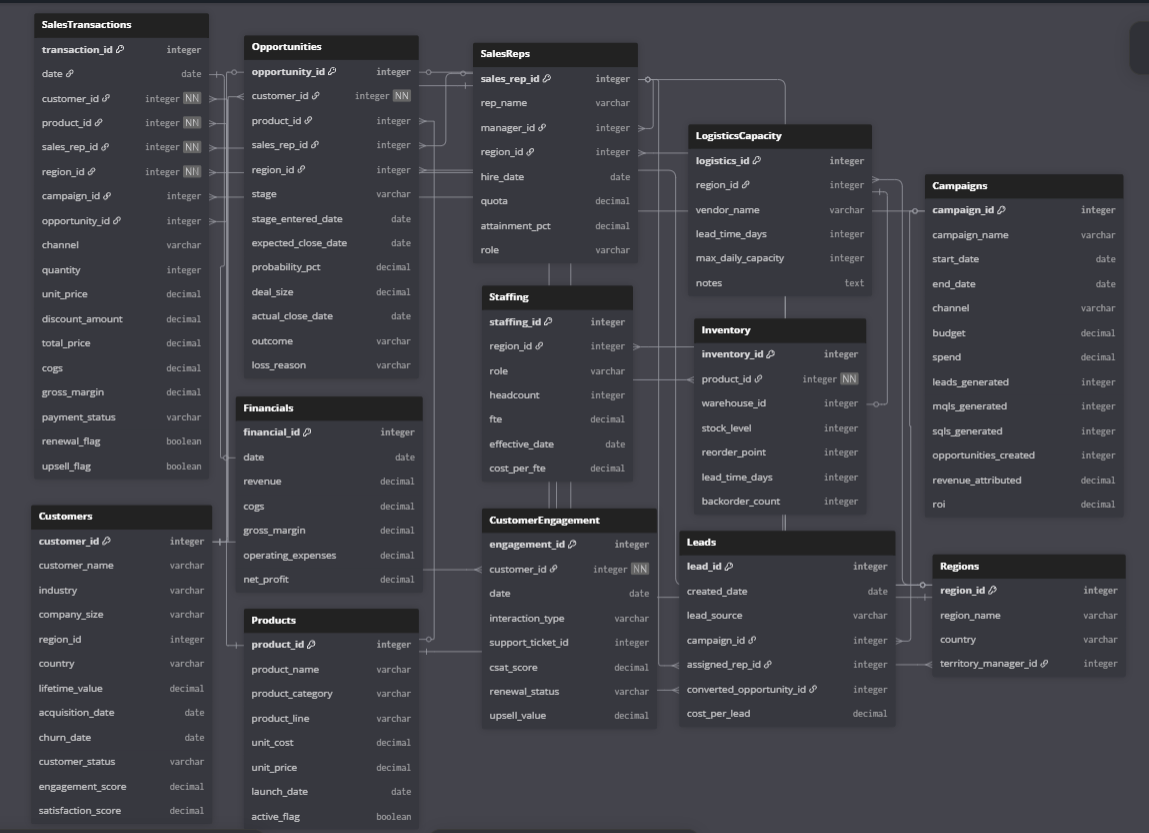
# Data Requirements

ContinuumAi does not require data for training since it leverages pretrained LLMs like GPT-5 or Deepseek-R1 for task completion. However, data will be needed in designing the architecture of the entire system. This will help us understand what kind of data we can expect from the target user, and hence design our agents accordingly.

|  |  |
| --- | --- |
|  |  |
| **Data Sources** | * No proprietary training dataset is required, as ContinuumAI leverages pretrained LLMs (e.g., GPT-5, Gemini 2.5). * System design is guided by the **Universal Sales Schema(\*schema provided at end for reference)**, which captures ~90% of possible user questions, KPIs, and insights across sales operations. * For testing and validating performance of agents:    + **Synthetic Data** will be generated to populate the Universal Schema tables (e.g., sales transactions, opportunities, campaigns, staffing).   + **Public Datasets** will also be used when appropriate, sourced from reliable repositories such as Kaggle, Github, HuggingFace. We have already selected some sources:     - <https://www.kaggle.com/datasets/vivek468/superstore-dataset-final>     - <https://www.kaggle.com/datasets/kyanyoga/sample-sales-data>     - <https://www.kaggle.com/datasets/mohammadtalib786/retail-sales-dataset>     - <https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting>     - <https://www.kaggle.com/datasets/heemalichaudhari/adidas-sales-dataset>     - <https://www.kaggle.com/datasets/thedevastator/unlock-profits-with-e-commerce-sales-data> |
| **Data Requirements** | * **Note:** The following requirements apply to the synthetic and open-source datasets that will be used for designing and testing ContinuumAI’s agents. * **Dataset Size:**   Flexible to support both small and large inputs. For **testing**, ~1,000 – 10,000 synthetic records distributed across schema tables are sufficient to simulate diverse scenarios. For **real user input**, even small datasets (e.g., ~500 rows of sales transactions) are fully supported for generating meaningful insights.   * **Checks for Synthetic Data:**   Synthetic datasets must closely mimic **real-world patterns** to ensure reliability during testing. Controlled anomalies (e.g., missing values, unusual discounts, seasonal sales spikes) will be introduced to make the data realistic. Before use, synthetic data will be validated by generating **plots, KPIs, and insights internally** to confirm that they follow expected business trends (e.g., seasonality, growth, regional differences). Any inconsistencies will be adjusted to maintain data quality while still reflecting real-world variability.   * **Diversity of Data:**   Synthetic data will cover a **wide range of sales domains** to ensure generalizability:   * + **Retail & Clothing** – seasonal demand, promotional discounts, high-volume transactions.   + **Manufacturing & B2B Sales** – bulk orders, long sales cycles, higher deal sizes.   + **E-commerce/Online Stores** – high-frequency transactions, digital campaigns, cart abandonment.   + **Subscription/Recurring Revenue Models** – renewals, churn, upsells, cross-sells.   This diversity ensures the system can handle the full spectrum of use cases defined in the universal schema. |
| **Model Requirements** | **Note:** This section outlines the requirements of the models that will be used to generate the insights and outputs requested by users.   * **Training Strategy**:   As mentioned earlier, the core system uses pretrained LLMs → no further training required.  However, for analytics asked by the user, lightweight models (e.g., linear regression, ARIMA, clustering, time-series forecasting) will be trained on user-provided datasets to generate the desired insights.   * **Frequency of Training**:   On-demand by the user, only when user data is ingested for BI tasks.   * **Accuracy Targets**:   Minimum 80% accuracy (or equivalent metric) on user-provided data where supervised labels exist.  Time-series forecasts evaluated with RMSE/MAE thresholds. **Optimization Methods:** To enhance the accuracy and quality of outputs, ContinuumAI’s analytical agents will experiment with different model hyperparameters and compare results. Depending on the use case (e.g., forecasting vs. classification), agents may adjust parameters such as learning rates, regularization strength, or window sizes for time-series. The goal is to identify the configuration that produces the most reliable and interpretable results for the user’s dataset in the limited time frame. **Validation and Evaluation:** Since datasets are provided by the user, model validation will follow a **train/test split strategy**:   * + **Training**: Models are trained on ~80% of the provided data.   + **Testing**: The remaining ~20% is reserved for validation.   + **Interpretation**: If performance on the test set is acceptable, the model is deemed reliable for forecasting, classification, or other predictive tasks.   + **Evaluation Metrics:** Accuracy, F1 score, RMSE, MAE, R². |

### 

### **\*Universal Sales Schema – Overview**



The **Universal Schema** is a consolidated data model designed to cover ~80–90% of sales-related business questions and KPIs across different roles. While users will typically provide only a subset of these tables (2–4 depending on their focus), the schema ensures that all major **use cases (UC-001 to UC-020)** can be supported.

### **Table Descriptions**

* **SalesTransactions** Core fact table capturing every sales transaction with details on customer, product, rep, region, pricing, discounts, and margins.  
   **Use Cases:** UC-001 Revenue Drivers, UC-002 Quota Tracking, UC-005 Trend Drilldown, UC-006 Forecasting.  
   **KPIs:** Revenue, Gross Margin, Discount Impact, Sales by Region/Product.
* **Opportunities** Pipeline and deal tracking table with stages, probabilities, deal size, and outcomes.  
   **Use Cases:** UC-003 Rep Benchmarking, UC-004 Pipeline Gap Closure.  
   **KPIs:** Win Rate, Average Deal Size, Pipeline Coverage.
* **CustomerEngagement** Records customer interactions, support tickets, satisfaction, upsell/renewal status.  
   **Use Cases:** UC-010 Churn Forecasting, UC-011 At-Risk Accounts, UC-012 Cross-Sell, UC-013 Customer Health, UC-014 Renewal Prediction.  
   **KPIs:** Churn Risk, Engagement Scores, Renewal Rates.
* **Financials** High-level financial metrics for reconciliation and profitability tracking.  
   **Use Cases:** UC-001 Revenue Drivers, UC-017 Staffing vs. Demand.  
   **KPIs:** Revenue, COGS, Net Profit, Operating Margin.
* **Inventory** Tracks product stock levels, reorder points, and backorders across warehouses.  
   **Use Cases:** UC-008 Product Demand Forecast, UC-016 Inventory Risk, UC-018 Supply Chain Stress Test.  
   **KPIs:** Stockouts, Demand vs. Supply, Inventory Turnover.
* **Customers** Master table for customer profiles including industry, size, lifetime value, churn/acquisition dates.  
   **Use Cases:** UC-010 Churn Forecast, UC-013 Health Dashboard, UC-015 Cohort Analysis.  
   **KPIs:** Customer Lifetime Value, Retention, Segmentation.
* **Products** Catalog of products with categories, unit costs, prices, and active status.  
   **Use Cases:** UC-007 What-If Pricing, UC-008 Demand Forecast.  
   **KPIs:** Product Profitability, Category Performance.
* **SalesReps** Information on sales reps, quotas, attainment, and hierarchy.  
   **Use Cases:** UC-002 Quota Tracking, UC-003 Rep Benchmarking.  
   **KPIs:** Quota Attainment, Revenue per Rep, Team Productivity.
* **Regions** Geographic breakdown of sales territories.  
   **Use Cases:** UC-005 Trend Drilldown, UC-006 Territory Forecasting, UC-018 Supply Chain Stress Test.  
   **KPIs:** Regional Sales, Territory Growth, Manager Performance.
* **Campaigns** Marketing campaign data including spend, leads, conversions, and attributed revenue.  
   **Use Cases:** UC-009 Campaign Impact Forecast, UC-019 ROI Comparison, UC-020 Campaign Alerts.  
   **KPIs:** Campaign ROI, Lead-to-Revenue Conversion, Attribution Metrics.
* **Leads** Core for marketing → sales attribution and funnel tracking.  
   **Use Cases:** UC-009 (Campaign Impact Forecast), UC-019 (Lead Source ROI), UC-020 (Campaign Underperformance Alerts).  
   **KPIs/Insights:** Lead volume, Cost Per Lead (CPL), Lead → Opportunity conversion rate, campaign attribution, MQL→SQL efficiency.
* **Staffing** Captures headcount/FTE by region and role for capacity planning.  
   **Use Cases:** UC-017 (Staffing vs. Sales Demand).  
   **KPIs/Insights:** FTE by region/role, utilization vs forecasted demand, staffing gaps, cost-per-FTE and suggested staffing adjustments.
* **LogisticsCapacity** Models vendor/warehouse capacity and lead-times for supply-chain scenarios and stress tests.  
   **Use Cases:** UC-018 (Supply Chain Stress Test), UC-016 (Inventory Risk Alerts — supplementing Inventory).  
   **KPIs/Insights:** Vendor lead time, max daily throughput, warehouse capacity utilization, constraints for stress-test simulations.

# Non-functional Requirements / Quality Attributes

|  |  |
| --- | --- |
| **Sr#** | **Requirements** |
| 1 | The system will process and display visualizations for 90% of natural language queries on datasets under 500 MB within 10 seconds. |
| 2 | A 100 MB CSV file will be ingested, processed, and made ready for analysis in under 60 seconds. |
| 3 | After a 15-minute tutorial, new users must achieve an 80% first-attempt success rate when generating a sales trend visualization without assistance. |
| 4 | The system must support 20 concurrent users with no more than a 20% degradation in average query response time compared to a single user. |
| 5 | Exporting any chart to a PNG file or data table to a CSV file must complete within 5 seconds. |

# Security Requirements

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr#** | **Security Risks** | **Potential Losses** | **Controls** |
| 1 | Broken Access Control | Confidential data exposure, Data integrity compromise,  Legal and compliance issues | Session Isolation,  Access Control Validation,  Token Management |
| 2 | Cryptographic Failures | Confidential Data Exposure,  Data Integrity Compromise,  Regulatory Penalties | Encrypt Data at Rest and Transit,  Password Hashing,  Key Management |
| 3 | Injection | Confidential data exposure, Data integrity compromise,  System Compromise | Input Validation and Sanitisation,  Parameterized Queries for SQL |
| 4 | Unrestricted Resource Consumption | Denial of Service (DoS), Increased Operational Costs | Rate Limiting,  Resource Quotas (for network bandwidth etc.) |
| 5 | Server Side Request Forgery (SSRF) | Data Exposure,  Denial of Service (DoS) | Input Validation and Sanitization,  Allowlists for URL Origins |

# Security Engineer

|  |  |
| --- | --- |
| **Name of the Security Engineer** | Muhammad Nafees |

# Use of Generative AI

<Mention here how generative AI was used in preparation of this artifact.>

* Section 3: Used for brainstorming and refining use cases in conjunction with primary research conducted with industry experts
* Section 8:Used a little inspiration in regards to standard non functional requirements

# Who Did What?

|  |  |
| --- | --- |
| **Name of the Team Member** | **Tasks done** |
| Ali Faizan | Sections 1, 2, 7 |
| Muhammad Bazaf Shakeel | Section 2, 4, 5 |
| Umer Raja | Sections 1, 2, 3, 11 |
| Mustufa | Section 4,8,11 |
| Nafees | Section 6, 9 , 10 |

# Review checklist

|  |  |
| --- | --- |
| **Section** **Title** | **Reviewer Name(s)** |
| Data Requirements | Nafees |
| Use Cases | Mustufa |
| Class Diagrams | Bazaf |
| State Diagrams | Ali Faizan |
| Sequence Diagrams | Umer Raja |