# LA Clippers Business Insights & Analytics Data Challenge

## 1. Data Dictionary

|  |  |  |  |
| --- | --- | --- | --- |
| Sheet Name | Description | Key Columns | Primary Use |
| Retail+F&B 4.24 | Item-level transaction data (April 24 game) | TransactionId, CustomerAccount, ItemId, ProductName, Category, BusinessVertical, Quantity, NetAmount, DiscountAmount, Store, TransactionDateTime | Analyze retail + F&B spend and item categories |
| Retail+F&B 4.26 | Item-level transaction data (April 26 game) | Same as above | Combine for multi-game revenue comparison |
| Retail+F&B 5.1 | Item-level transaction data (May 1 game) | Same as above | Detect shift in item mix and basket size |
| EntryScans | Fan ticket scan data (entry logs) | EventName, RedeemingFanNBAId, RedemptionDateTimeUTC, DeviceLocation, RedemptionMethod, SectionName | Analyze arrival times & entrance patterns |
| CustomerIDs | Fan lookup table linking CustomerAccount ↔ NBA ID | NBAId, CustomerAccount, DistanceToIntuitDome, AppCreatedDatetime | Merge fan info to spending and entry records |
| StoreEntries 4.24 | In-store entry logs (Apr 24) | eventid, StoreEnterDate, StoreEnterDateTime, NBAId, CheckpointName, accessmethodname | Calculate store conversion and visit frequency |
| StoreEntries 4.26 | In-store entries (Apr 26) | Same schema | Compare store traffic across games |
| StoreEntries 5.1 | In-store entries (May 1) | Same schema | Assess variation in store traffic patterns |
| Arena Pricing Map | Image of arena zones and section pricing | N/A | Map section names to price tiers / zones |
| StoreNames | Store ID ↔ Friendly Name lookup | StoreID, StoreName | Join to Retail+F&B and StoreEntries |

## 2. Product Requirements Document (PRD)

### Objective

Identify two actionable insights and one recommendation to enhance fan experience and optimize in-arena revenue through Retail + F&B data.

### Business Problem

Understand how fan arrival, store activity, and purchase trends connect to optimize operations and fan engagement.

### Success Metrics

Arrival timing distribution, conversion rate, revenue per fan, basket size, zone performance.

### Data Relationships

CustomerIDs ↔ Retail+F&B ↔ StoreEntries ↔ EntryScans ↔ Arena Pricing Map.

### Analytical Questions

Do early arrivals spend more? Which entrances convert best? Which zones yield higher spend? When do fans buy merch vs F&B? Do local fans spend more?

### Analytical Approach

Data ingestion → Cleaning → Feature Engineering → EDA → Insight Selection → Recommendation Development → Deliverables.

### Deliverables

One-page summary, 3-slide deck, Analysis notebook/Excel, ≤3-min video.

### Tools

Python (pandas, matplotlib), Jupyter, Excel, PowerPoint, Loom/Zoom.

## 3. Action Plan

### Step 1: Data Exploration & Setup (DONE)

- Load Excel sheets into DataFrames

- Inspect schemas and missing values

- Document column meanings in Data Dictionary

## 🧹 ****Step 2: Data Cleaning & Standardization****

### ****2.1 — Column Name & Type Normalization****

* Convert all headers to snake\_case (or consistent camelCase).
* Strip whitespace and special characters (+, &, etc.).
* Standardize data types:
  + Dates → datetime64
  + IDs → string
  + Amounts → float
* Verify consistent schemas across date-based sheets (Retail+F&B 4.24, 4.26, 5.1).

### ****2.2 — Missing & Duplicate Handling****

* Generate missing-value summary per column (thresholds ≥ 50% flagged).
* Impute or drop based on context (e.g., missing DiscountAmount → 0).
* Remove duplicates by primary keys (TransactionId, CustomerAccount, etc.) and log removed counts.

### ****2.3 — ID & Key Standardization****

* Clean IDs:
  + Trim whitespace and zero-pad where needed.
  + Ensure all Store IDs match entries in StoreNames.
  + Validate CustomerAccount pattern (C#########).
* Create surrogate keys if needed for consistency across date-suffixed sheets.

### ****2.4 — Timestamp Alignment****

* Convert all timestamps to a consistent format and timezone (PST or UTC).
* Standardize column names (TransactionDateTime, EntryDateTime, etc.)
* Extract useful derived columns (date, hour, day\_of\_week).

### ****2.5 — Value Normalization****

* Standardize categorical values (e.g., “Food and Beverage” vs “Food & Beverage”).
* Ensure currency/amount fields use a uniform numeric scale.
* Validate logical constraints (Quantity ≥ 1, NetAmount ≥ 0).

### ****2.6 — Consistency Checks & Audit Logs****

* Validate referential integrity (StoreEntries ↔ StoreNames, CustomerAccount ↔ CustomerIDs).
* Log every transformation to /logs/cleaning\_step2.log.
* Save cleaned versions to /data\_interim/.

### ****2.7 — Final Consolidation****

* Merge all “Retail+F&B” sheets into one master table with a SourceDate column.
* Do the same for “StoreEntries.”
* Produce /docs/step2\_summary.md with before/after stats and major cleaning actions.

## ****Step 3 — Feature Engineering & Data Integration****

### ****3.1 — Data Integration Setup****

* Load both master datasets:
  + retail\_fnb\_master.parquet
  + store\_entries\_master.parquet
* Verify column types and join keys (store\_id, customer\_account, nba\_id)
* Conduct quick join tests to confirm relationship coverage
* Produce /docs/step3\_01\_integration\_setup.md with join readiness summary

### ****3.2 — Master Join & Enrichment****

* Perform key joins to enrich transaction data with:
  + **Store entries** (match on store\_id + source\_date)
  + **CustomerIDs** (map via customer\_account or nba\_id)
* Handle non-matches and duplicates carefully
* Save joined table → /data\_interim/step3\_02\_enriched\_master.parquet
* Document join coverage and unmatched counts → /docs/step3\_02\_join\_summary.md

### ****3.3 — Feature Derivation: Customer Behavior****

* Engineer customer-centric features:
  + total\_transactions
  + avg\_spend\_per\_visit
  + unique\_store\_visits
  + repeat\_visit\_flag
  + avg\_time\_between\_entries
* Aggregate at customer\_account or nba\_id level
* Save to /data\_interim/step3\_03\_customer\_features.parquet

### ****3.4 — Feature Derivation: Store & Event Performance****

* Create store-level and event-level metrics:
  + avg\_transaction\_value
  + transactions\_per\_game
  + peak\_hour\_sales\_ratio
  + entry\_to\_purchase\_conversion\_rate
* Aggregate at store\_id + source\_date
* Save to /data\_interim/step3\_04\_store\_event\_features.parquet
* Document KPIs → /docs/step3\_04\_store\_metrics.md

### ****3.5 — Feature Scaling & Encoding****

* Normalize numerical features (z-score or min-max)
* Encode categorical variables (e.g., store zone, day\_of\_week)
* Handle missing derived metrics (impute 0 or mean)
* Produce training-ready dataset → /data\_processed/step3\_05\_feature\_ready.parquet
* Log transformations → /logs/step3\_05\_feature\_scaling.log

### ****3.6 — Feature Validation & Export****

* Validate all engineered features:
  + No NaNs in model-ready fields
  + Expected ranges confirmed (e.g., ratios ≤ 1)
  + Schema consistent with modeling specs
* Compute feature correlation matrix & detect multicollinearity
* Save documentation:
  + /docs/step3\_06\_feature\_validation.md
  + /docs/step3\_06\_feature\_dictionary.md
* Final output: /data\_processed/feature\_engineered\_master.parquet

### ✅ ****Deliverables at the End of Step 3****

* Integrated and feature-rich master dataset
* Feature dictionary and validation reports
* Correlation & data-quality diagnostics
* Ready input for modeling and visualization (Step 4)

### Step 4: Exploratory Data Analysis (EDA)

- Visualize arrival patterns

- Analyze spend by zone, time, and entrance

- Identify consistent patterns across games

### Step 5: Insight Development

- Select top 2 insights based on impact and consistency

- Quantify KPI improvement (e.g., +X% revenue)

### Step 6: Recommendation Formulation

- Link strongest insight to an actionable operational change

- Estimate pilot impact and define success metric

### Step 7: Deliverable Creation

- Write 1-page summary

- Design 3-slide deck

- Record ≤3-minute video walkthrough

- Prepare clean analysis notebook

### Step 8: QA & Submission

- Cross-verify figures between notebook and deck

- Ensure consistent file naming

- Email deliverables before Oct 22 EOD