Dynamic Quest Offering

1. Problem Understanding & ML Framing

1.Problem Statement

Cruise lines offer a wide array of ancillary services such as specialty dining, shore excursions, spa treatments, and onboard activities. Traditionally, marketing these services has been generic, leading to missed revenue opportunities and suboptimal guest experiences.

The challenge is to develop a dynamic, personalized recommendation engine that suggests ancillary services tailored to each guest. The recommendations must be context-aware, taking into account:

- Guest profile: demographics, loyalty tier, party type, and persona.
- Guest behavior: past purchase history, spending patterns, and item preferences.
- Sailing context: cruise itinerary, ship type, season, region, and available onboard amenities.
- Special occasions: birthdays, anniversaries, or other events that can influence preferences.

The goal is to increase high-margin onboard revenue while enhancing guest satisfaction by providing a digital concierge experience, making every vacation more relevant and memorable for the guest.

2.ML Framing

Given the problem, the recommendation task is framed as a hybrid recommendation/ranking problem:

1.Cold-start guests (no history):

Use feature-based recommendations leveraging guest profile, sailing context, season, region, and item popularity.

2.Sparse-history guests (limited history):

Use user-based collaborative filtering within guest clusters and similarity groups.

Incorporate item correlations (e.g., Dining ↔ Entertainment) and persona-based boosts to improve relevance.

3.Rich-history guests (extensive history):

Use user-based CF with scoring enhancements: consider purchase frequency, spend, loyalty tier weighting, and special occasion boosts.

Integrate contextual filters for itinerary, ship amenities, and seasonal availability.

3. Key ML Task Types:

Ranking / Recommendation: Generate a top-N list of ancillary items personalized for each guest.

Feature Engineering: Encode guest profiles, purchase behavior, itinerary, seasonality, and item attributes into numeric or categorical features for scoring.

Hybrid Approach: Combines collaborative filtering, feature-based scoring, and business-driven boosts (persona, loyalty, special occasions).

4.Business-Oriented Objective Function:

Maximize predicted revenue from top-N recommended items while ensuring relevance and personalization.

Ensure recommendations align with premium guest profiles and special occasions to increase upsell potential.

2. Data Exploration Report (EDA)

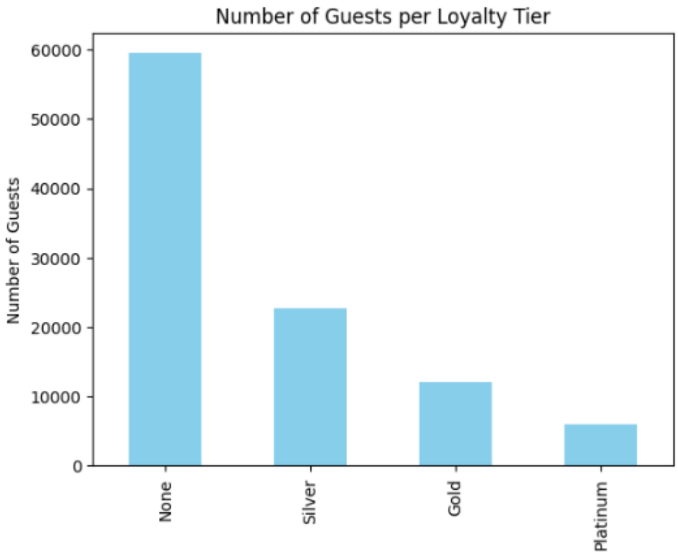
The dataset for this recommendation engine includes multiple sources:

- Guests: Guest demographics, loyalty tier, age, gender, country, and party type.
- Items: Ancillary offerings including dining, spa, excursions, and entertainment items with category, price, and original price.
- Purchases: Historical purchases linked to guests and sailings.
- Sailings & Itineraries: Cruise itineraries, ship types, regions, quarters, and duration.
- Bookings: Additional context including special occasions.
- Itinerary: Data about itinerary of sailing
- Ships: Data about ships offering sailing and ship type with amenities provided

2.1 Guest Profile Insights

- Loyalty Tiers: Majority of guests fall under "None" and Silver, while a smaller segment are Gold/Platinum. Higher tiers correspond to higher spend potential.

```
loyalty_tier
None        59419
Silver      22653
Gold        12018
Platinum     5910
Name: count, dtype: int64
```



- Age Distribution: Guests range across multiple age buckets, with clusters such as young explorers, families, and mature travelers.
- Party Types: Singles, couples, and families are well-represented, influencing preferences for activities and amenities.
- Missing Values: Minimal in demographic fields; loyalty_tier occasionally missing, handled by filling with "None"

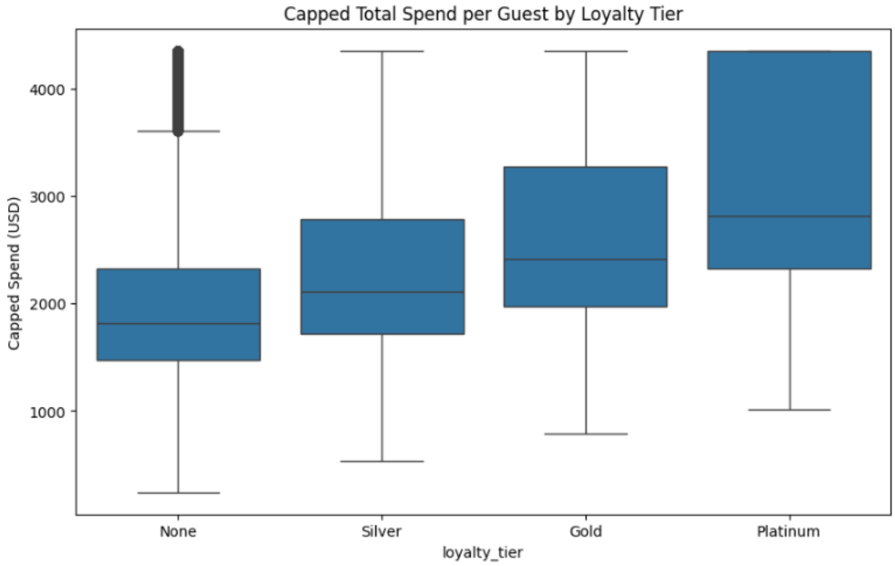
2.2 Purchase Behaviour

Spend Patterns:

High-margin services such as specialty dining, spa, and premium excursions dominate revenue for rich-history guests.

Average spend per guest shows right-skewed distribution (few high-spending guests).

	loyalty_tier	count	sum	mean	median
0	Gold	7102	1.905717e+07	2683.352174	2414.17
1	None	25589	5.197967e+07	2031.328555	1819.42
2	Platinum	3967	1.211902e+07	3054.957916	2814.00
3	Silver	11901	2.833168e+07	2380.613191	2110.36



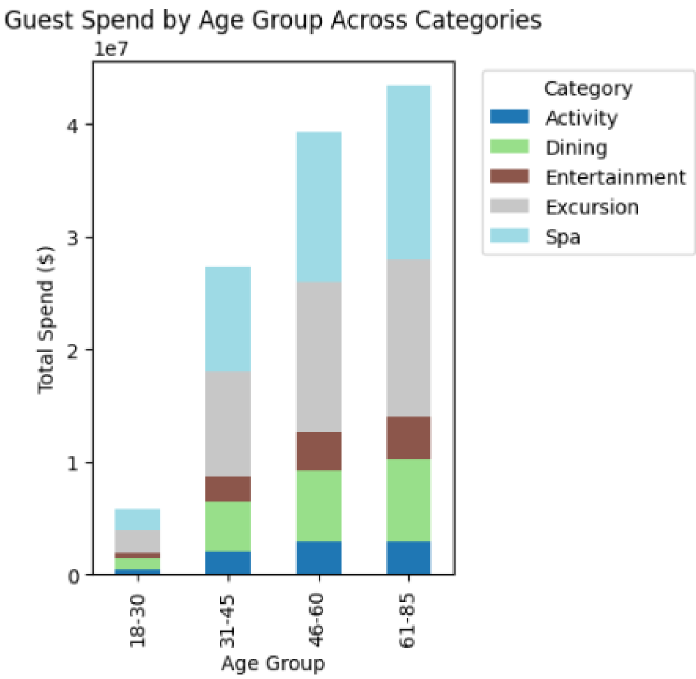
Tier Based Spending

Note: The Spending Behaviour is having correlation between Loyalty Tiers, As Platinum is the high spend group

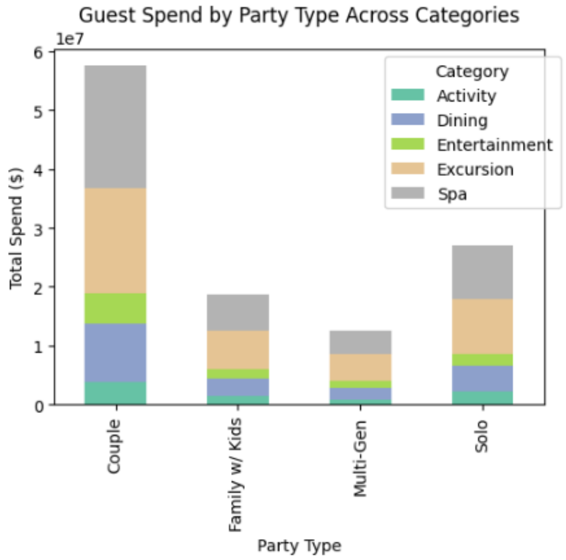
**** Outlier from purchases where capped using IQR capping*

For Cold start guests Age/Party type based recommendation can be suggested. As we don't have any previous histories for these kind of users , recommendations can be based on purchases done by similar Age groups,Party type

<u>Age Category Spend</u>



Party Based Spend



Target Personas:

- *Luxury / Premium Mature → Couples, mid to older age → maximize revenue with high-margin bundles.*
- *Adventure / Wellness Seekers → Young guests, Solo travelers → focus on Excursion/Activity/Spa.*
- *Family / Group → 31-45 age, Families → bundle activities + dining to improve cross-sell.*

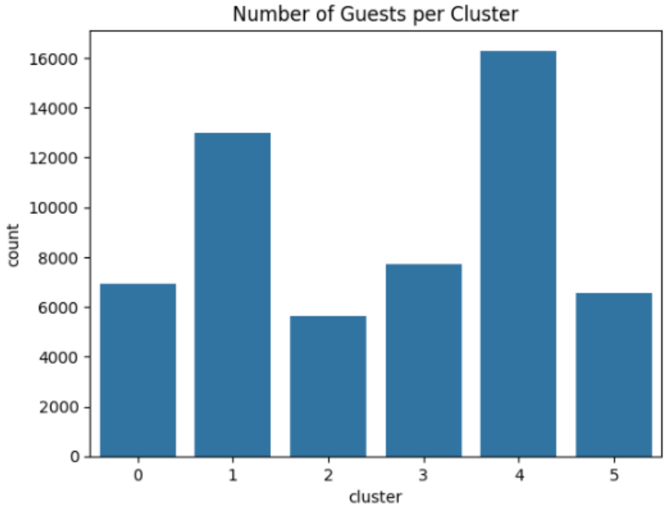
Seasonal / Contextual Relevance: *Combine spend patterns with itinerary, ship type, and season to enhance recommendation accuracy.*

Cold Start & Sparse Guests: *Recommend top popular items by* ***region + season*** *to capture new revenue.*

Market Basket Analysis: *Guides* ***bundling strategy*** *for cross-selling high-margin items while avoiding negatively correlated bundles.*

Guests Clustering based on similar Behaviour (Demographics-Age, Party type, Spending, Ship type chosen)

Cluster	Guest Type
0	Luxury Couples
1	Family Adventure Seekers
2	Solo Entertainment Travelers
3	Mid-range Couples & Spa Lovers
4	Young Explorer Families
5	Premium Mature Travelers

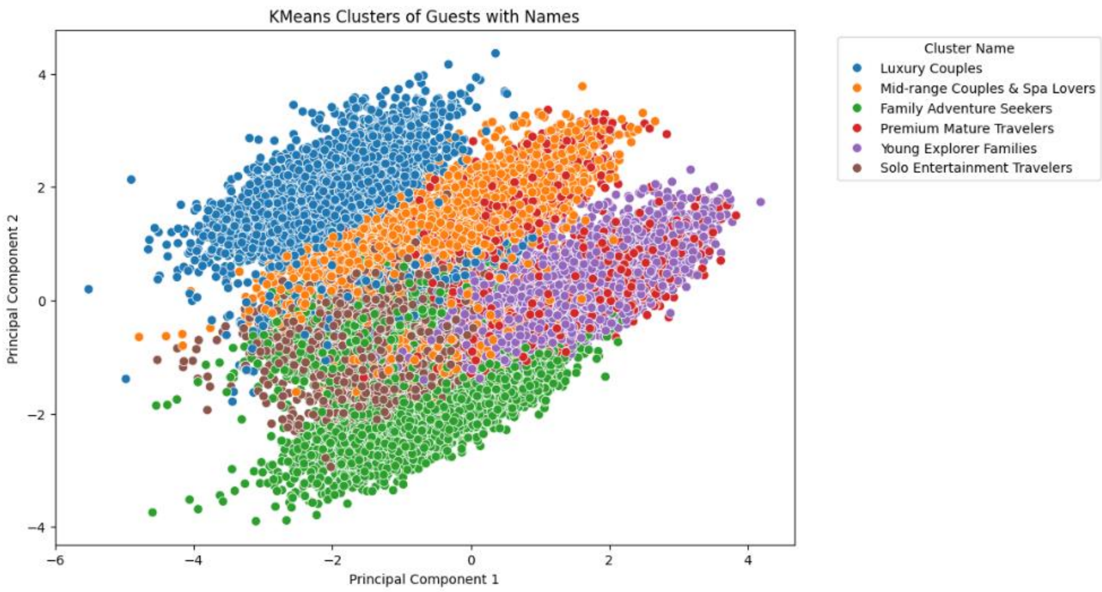


category cluster	Activity	Dining	Entertainment	Excursion	Spa
0	0.169371	0.273969	0.102174	0.225179	0.229307
1	0.185364	0.263805	0.098946	0.215907	0.235978
2	0.165956	0.262702	0.111609	0.234211	0.225521
3	0.151026	0.277040	0.114477	0.220085	0.237372
4	0.145275	0.292016	0.111898	0.199414	0.251397
5	0.147208	0.284283	0.109914	0.199743	0.258852

*** Most of the clusters inclined to Dining

	cluster	item_id	item_name	count
941	0	I0942	Orchestra Dinner Show	368
969	0	I0970	Dance Spectacular	364
1870	1	I0871	Rock Climbing Masterclass	778
1935	1	I0936	Pilates Class	765
2940	2	I0941	Exclusive Stand-up Comedy Night	323
2948	2	I0949	Exclusive Stand-up Comedy Night	321
3941	3	I0942	Orchestra Dinner Show	490
3940	3	I0941	Exclusive Stand-up Comedy Night	453
4948	4	I0949	Exclusive Stand-up Comedy Night	1042
4984	4	I0985	Acrobatic Dinner Show	1026
5987	5	I0988	Exclusive Stand-up Comedy Night	441
5941	5	I0942	Orchestra Dinner Show	437

Top 2 Items preferred by clusters



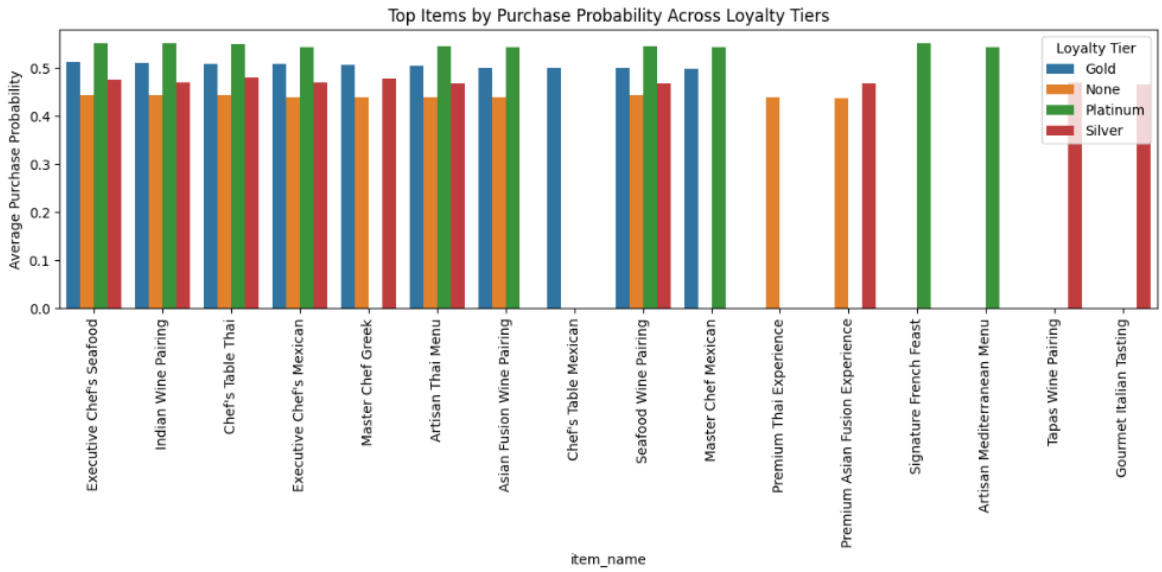
PCA is used here for dimensionality reduction

Inference

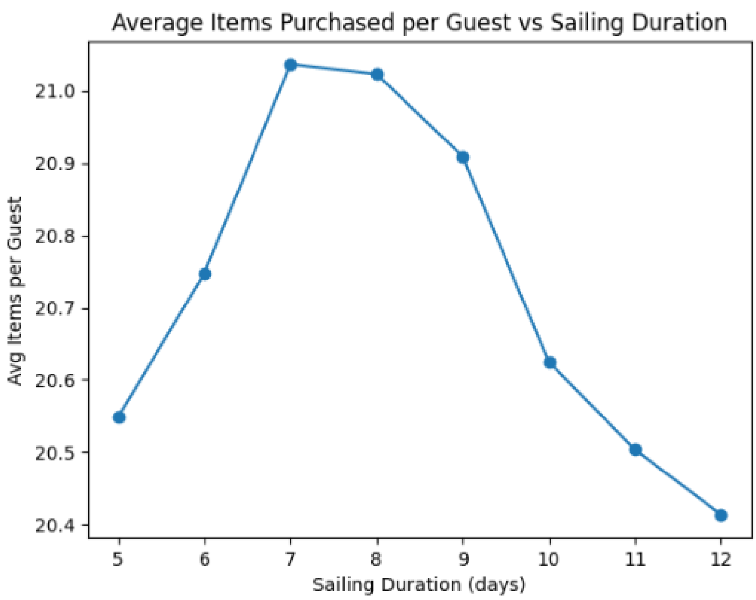
- ✓ *Luxury Couples and Premium Mature Travelers may represent high-value customers who respond well to premium or exclusive offers.*
- ✓ *Family Adventure Seekers and Young Explorer Families might prefer packages with activities and group-friendly pricing.*
- ✓ *Solo Entertainment Travelers likely value experiences and convenience over luxury amenities*

Purchase Frequency & Preferred Items by Loyalty tiers

- ✓ Items like Dining upgrades and Spa treatments are consistently purchased by loyal guests, confirming the importance of loyalty-aware recommendations.
- ✓ Guests purchasing Dining upgrades often purchase Entertainment but rarely choose Excursions simultaneously.



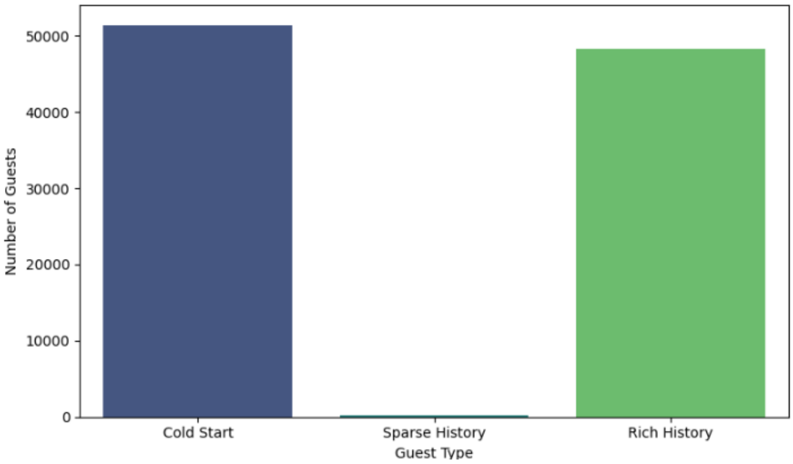
Sailing Duration Influence on purchases



- ✓ Guests tend to purchase more items on medium-length cruises (7–8 days).
- ✓ For shorter or longer cruises, the average purchases per guest are lower — possibly because:
    - Shorter trips have less time/opportunity for purchases.
    - Longer trips may involve guests pacing their spending or reaching saturation.

When booking for duration comes for 7-8 days, its better to recommend the popular products across their clusters. (see clusters above)

Purchase Frequency: Some guests have sparse histories, while a few exhibit rich, repeated purchases. This supports hybrid recommendation strategies



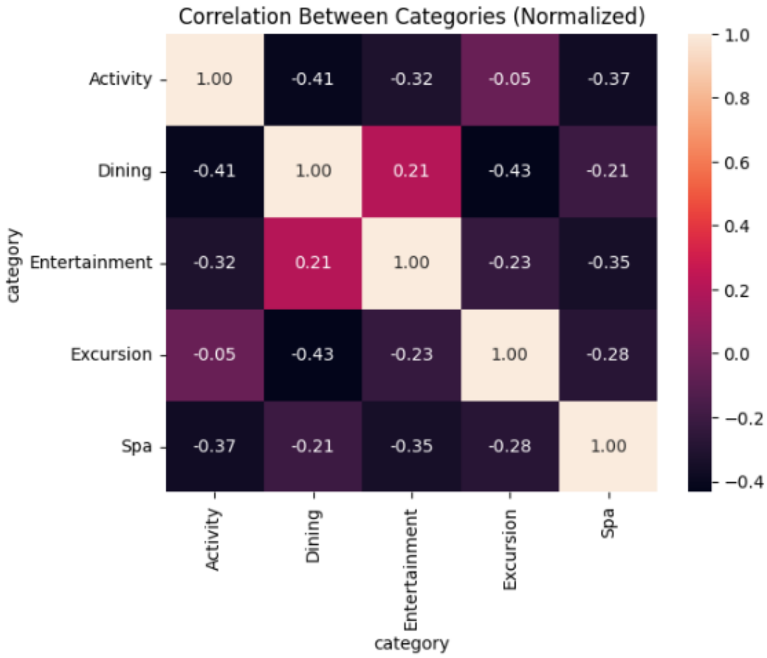
Majority are **Cold start** users who does not have any purchase histories

Special Occasion purchases

	special_occasion	item_name	category	num_purchases
940	Anniversary	Exclusive Stand-up Comedy Night #01	Entertainment	152
970	Anniversary	Stand-up Comedy Show #31	Entertainment	151
948	Anniversary	Exclusive Stand-up Comedy Night #09	Entertainment	143
1984	Birthday	Acrobatic Dinner Show #45	Entertainment	178
1948	Birthday	Exclusive Stand-up Comedy Night #09	Entertainment	174
1940	Birthday	Exclusive Stand-up Comedy Night #01	Entertainment	173
2882	Graduation	Beginner's Photography #35	Activity	37
2939	Graduation	Orchestra Dinner Show #02	Entertainment	36
2951	Graduation	Premium Jazz Event #14	Entertainment	35
3985	Honeymoon	Exclusive Stand-up Comedy Night #48	Entertainment	116
3959	Honeymoon	Orchestra Show #22	Entertainment	113
3968	Honeymoon	Stand-up Comedy Show #31	Entertainment	113
4939	None	Orchestra Dinner Show #02	Entertainment	2928
4985	None	Exclusive Stand-up Comedy Night #48	Entertainment	2882
4946	None	Exclusive Stand-up Comedy Night #09	Entertainment	2847

This can be used for used based collaborative recommendation, on special occasions, these top items can be suggested across similar users with matching demographic patterns

Item Popularity across itineraries

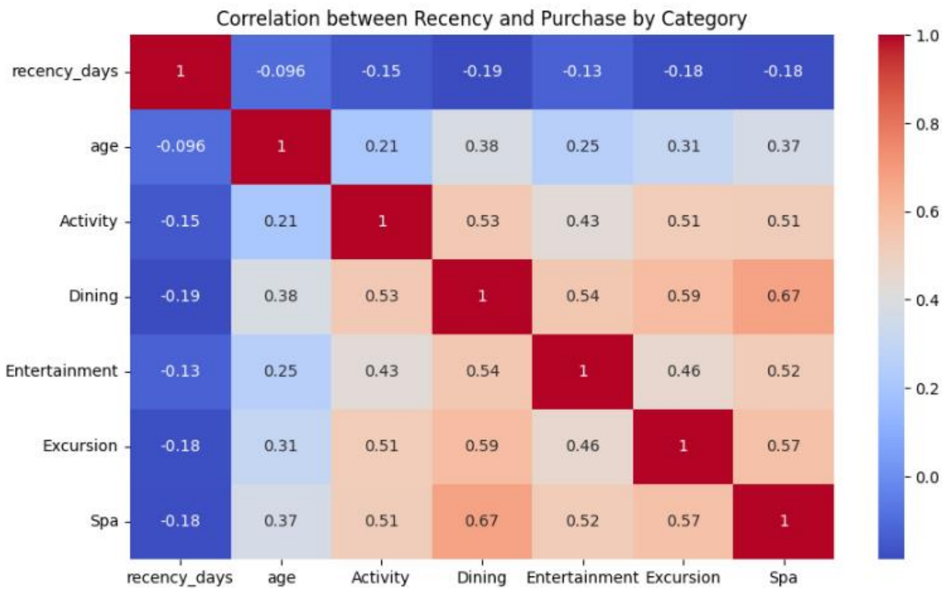


- ✓ Dining and Entertainment show mild positive correlation (r = 0.21). Guests who enjoy dining often purchase entertainment.
- ✓ Dining and Excursion show negative correlation (r = -0.43). Guests preferring excursions less likely to upgrade dining.
- ✓ Activity and Spa show negative correlation (r = -0.37), reflecting distinct guest personas: adventure-focused vs. relaxation-focused.

.

Recency Correlation:

Correlation between recency_days and purchase category is very **low**, suggesting that how recently a guest travelled does not strongly influence which category they purchase.



Sailing & Itinerary Insights

Regions: Mediterranean and Caribbean dominate, consistently showing highest ancillary spend, particularly for Dining and Spa. Seasonal availability affects recommendations, e.g., shore excursions limited in Winter routes.

Mediterranean and Caribbean itineraries consistently dominate in every season — both regions show the highest spending on Dining and Spa services, suggesting a premium customer base focused on comfort and onboard experiences.

Excursion and Activity bookings are notably higher in Caribbean and Mediterranean cruises during Spring and Summer, reflecting the popularity of warm-weather adventure and sightseeing packages.

Northern Europe itineraries show moderate participation across categories, with balanced interest in Dining and Entertainment but lower engagement in Spa and Excursion activities, likely due to shorter cruise durations.

In Asia, Dining and Entertainment purchases rise during Spring and Fall, indicating that Asian routes may appeal to guests seeking cultural experiences over adventure-based excursions.

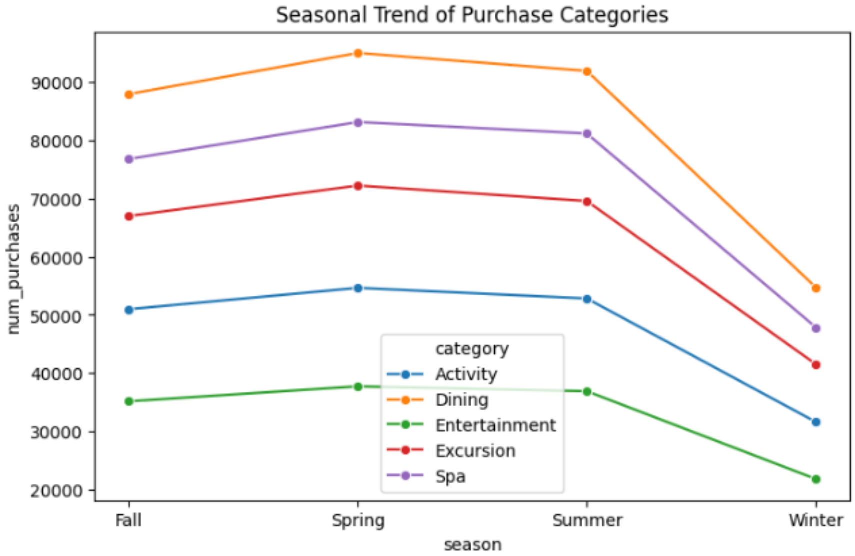
Australia/New Zealand shows the lowest total ancillary spend across all seasons, likely due to smaller ship sizes or fewer onboard revenue opportunities.

category		Activity	Dining	Entertainment	Excursion	Spa
season	region					
Fall	Alaska	4401.0	7722.0	3082.0	5859.0	6541.0
	Asia	5904.0	10223.0	4085.0	7978.0	8849.0
	Australia/NZ	3511.0	6314.0	2560.0	4753.0	5519.0
	Caribbean	14872.0	24978.0	9943.0	19166.0	21745.0
	Mediterranean	14229.0	24453.0	9815.0	18554.0	21634.0
	Northern Europe	8058.0	14286.0	5655.0	10691.0	12521.0
Spring	Alaska	4711.0	8403.0	3412.0	6259.0	7016.0
	Asia	6460.0	11236.0	4528.0	8560.0	9979.0
	Australia/NZ	3283.0	5914.0	2318.0	4374.0	5033.0
	Caribbean	15367.0	26329.0	10450.0	20194.0	23126.0
	Mediterranean	16382.0	28385.0	11205.0	21531.0	24856.0
	Northern Europe	8444.0	14765.0	5804.0	11320.0	13147.0
Summer	Alaska	6824.0	12341.0	5032.0	9213.0	10694.0
	Asia	4782.0	8312.0	3383.0	6325.0	7503.0
	Australia/NZ	3517.0	5991.0	2487.0	4631.0	5251.0
	Caribbean	12486.0	21366.0	8586.0	16238.0	18723.0
	Mediterranean	15725.0	27413.0	10873.0	20725.0	24087.0
	Northern Europe	9488.0	16527.0	6517.0	12446.0	14932.0
Winter	Asia	6291.0	10745.0	4294.0	8095.0	9272.0
	Australia/NZ	3878.0	6701.0	2701.0	5136.0	5811.0
	Caribbean	13469.0	23022.0	9105.0	17452.0	20229.0
	Mediterranean	4919.0	8890.0	3579.0	6794.0	7898.0
	Northern Europe	3041.0	5394.0	2105.0	4058.0	4666.0

Seasonality

- ✓ Caribbean & Mediterranean cruises see peak purchases in Spring/Summer.
- ✓ Northern Europe shows balanced interest in Dining and Entertainment, lower Spa and Excursion engagement.
- ✓ Asia sees Dining and Entertainment peak in Spring/Fall.

Duration & Ship Type: Longer cruises correlate with higher spend; premium ships attract luxury personas.



2.3 Feature Distribution and Handling Missing values-Data Wrangling

Feature Observations Handling.

- ✓ guest_id Unique, no missing
- ✓ loyalty_tier Some missing Fill with 'None'.
- ✓ Age value is Continuous,
- ✓ Some outliers Cap at 100+ or group into buckets.
- ✓ Party_type Categorical
- ✓ Spend Highly skewed Log transform possible for modeling.
- ✓ item_id / category All present
- ✓ season Derived from quarter Mapped via quarter to season.
- ✓ special_occasion Sparse Binary flag for boosting recommendations.

3. Feature Definition & Rationale

3.1 Objective

The primary goal of feature selection and engineering was to create a **rich guest-item interaction dataset** that captures behavioral, contextual, and business-driven factors influencing ancillary purchases (Dining, Spa, Excursion, Entertainment, and Activity). Each feature was designed to support **personalized and revenue-optimized recommendations**.

3.2 Feature Selection Process

1. **Correlation Analysis:**

    - Checked correlation between numeric features (recency_days, total_spend, num_purchases) and purchase categories.
    - Observed very **low correlation between recency and spend** → excluded from modeling but used for segmentation.
2. **Missing Value Handling:**
    - Imputed missing categorical attributes (e.g., loyalty_tier, party_type) with "None" or most frequent values.
    - Ensured no data loss during merging of guest, sailing, and purchase datasets.
3. **Outlier Treatment:**
    - Used **IQR-based capping** on spend and price to remove unrealistic transaction outliers.
4. **Encoding & Scaling:**
    - Applied **One-Hot Encoding** for categorical variables (e.g., party_type, season, region).
    - Used **Min-Max Scaling** for normalized distance and similarity computations (cosine similarity).
5. **Dimensionality Reduction:**
    - Implemented **PCA** (Principal Component Analysis) for behavioral clustering — improving computational efficiency and capturing principal guest traits.

3.3 Feature Engineering Process

Engineered Feature	Description	Purpose
season	Derived from sailing quarter (Q1→Winter, Q2→Spring, Q3→Summer, Q4→Fall	Models **seasonality of preferences** (e.g., Spa more in Winter, Excursion in Summer)
recency_days	Days since last purchase	Helps identify active vs dormant guests; weak correlation with purchase frequency noted, but useful for segmentation
spend	Total purchase amount per guest or per category	Used to measure spend intensity and assist in **persona identification**.
num_purchases	Count of purchases per guest	Differentiates cold-start, sparse, and rich-history guests.
guest_cluster	Derived via unsupervised clustering (k-means) on behavioral features (spend pattern, purchase diversity, recency)	Encodes latent guest behavior for **collaborative filtering and persona mapping**.

persona_map	Mapped clusters into interpretable personas (Luxury Couples, Family Adventure Seekers, etc.)	Enables rule-based boosting during recommendation.
itinerary_item_match	Joined itineraries with purchased items	Ensures recommendations are **contextually valid** (i.e., items available in the current sailing).
special_occasion_flag	Boolean derived from special_occasion column	Enhances personalization — boosts celebratory or premium recommendations
loyalty_boost	Derived from loyalty_tier (Gold/Platinum)	Used in the model to prioritize high-margin, exclusive recommendations.

4. Model Design Choices (Technical & Business Rationale)

4.1 Strategy Overview:

1. **Cold-Start Guests (No Purchase History):**
   - Recommendations rely on **popularity-based filtering**, taking into account **season, region, itinerary, and ship availability**.
   - Persona and loyalty-tier boosts are applied to highlight premium items for high-value guests.
   - Ensures relevant recommendations even when no historical behavior is available.
2. **Sparse / Rich History Guests:**
   - **Item-based collaborative filtering** is used, implemented via **precomputed item-item correlation scores**.
   - Each eligible item receives a **score** based on its correlation with items the guest has previously purchased.
   - Conceptually, this is equivalent to **cosine similarity** between item vectors, but computed via correlations to improve efficiency and interpretability in sparse datasets.
   - **Persona boost:** Certain guest segments (e.g., Luxury Couples, Premium Mature Travelers) have their scores adjusted to favor high-value items.

- **Loyalty-based revenue maximization:** Guests in higher loyalty tiers (Gold, Platinum) are prioritized with top-scoring items.

3. **Eligibility & Context Awareness:**
    - Only items available on the guest's ship and relevant to their sailing itinerary are considered.
    - Special occasions are used as an additional boost to recommend celebratory or exclusive items.
    - Season, region, and itinerary context ensures relevance to the guest's cruise experience.

4.2 Preprocessing & Feature Use:

- Feature engineering includes **guest demographics (age, party type, loyalty tier), sailing attributes (season, region, itinerary, ship type), and item metadata (category, price, spend, historical popularity)**.
- Purchase history is aggregated per guest to compute item correlations.
- Personas and loyalty tiers are used as **business-driven weighting factors**.

4.3 Business Alignment:

- This hybrid approach ensures recommendations are **personalized, context-aware, and revenue-maximizing**.
- Cold-start handling avoids generic recommendations.
- Correlation-based similarity (cosine-similarity–style) ensures that even sparse history guests receive relevant suggestions aligned with past behavior and business priorities.

5. Model Progression & Validation Strategy

5.1Baseline Model:

- Started with simple **popularity-based recommendations** for all guests, ranking items by overall purchase frequency.
- Evaluated using **NDCG@10**, focusing on how well top-10 recommended items matched actual guest purchases.
- Provided a reference point for improvement, especially for **cold-start guests**.

5.2 Enhanced Model (Final Pipeline):

- **Cold-start guests:** Used **region-, season-, and itinerary-based popularity** to recommend top items, as there is no purchase history.
- **Sparse-history guests:** Applied **feature-based similarity** and **item correlations**, leveraging guest demographics (age, party type, loyalty tier) and sailing context (itinerary, ship type, season).

- **Rich-history guests:** Applied **user-based collaborative filtering** with **cosine similarity across guest clusters**, combined with **persona-based boosts** for high-value segments.
- **Additional enhancements:**
    - Filtered recommendations using **eligible items per sailing/ship**.
    - Boosted recommendations for **special occasions** or **premium loyalty tiers**.
    - Recommended **top-10 items** to align with guest attention and business revenue goals.

5.3 Validation Strategy:

- Split historical data into **train and holdout sets**, ensuring guests in the holdout set mimic real test scenarios (cold, sparse, rich).
- Evaluated recommendations using **NDCG@10**, which measures:
    - Correct items appearing in the top ranks of recommendations.
    - Relevance of items weighted by position (higher ranks matter more).
- Iteratively tuned:
    - **Similarity weights**, **persona boost factors**, and **popularity thresholds**.
    - Top-N cutoffs to optimize **guest engagement and revenue impact**.

5.4 Outcome:

- The progression from simple popularity-based baseline to a **hybrid, context-aware recommendation pipeline** improved **NDCG@10**, ensuring more relevant items appear in the top-10 recommendations for all guest types.

6. Business Benefit & Evaluation Metrics

6.1 Business Benefit:

- The recommendation engine aims to **increase high-margin onboard revenue** by promoting relevant ancillary services such as dining, excursions, spa treatments, and entertainment.
- By personalizing offers based on guest **profile, loyalty, sailing context, and past behavior**, the model enhances **guest satisfaction**, encouraging repeat purchases and higher engagement.
- Special attention to **premium loyalty tiers and special occasions** ensures maximum revenue uplift for high-value customers.

6.2 Evaluation Metrics:

- **ML Metric: NDCG@10 (Normalized Discounted Cumulative Gain at 10)**

    - Measures how well the **top-10 recommended items** align with actual guest purchases.
    - Rewards items appearing at higher ranks in the recommendation list.
    - Suitable for the business objective since **only top-ranked recommendations are likely to drive actual purchases**.
- **Business Metric:**
    - **Revenue per guest / uplift in ancillary spend**: Tracks the financial impact of delivering personalized recommendations.
    - Helps quantify the **monetary benefit of improved recommendation relevance**.

6.3 Rationale for Metric Choice:

- **NDCG@10** balances **relevance and rank importance**, aligning with guest behavior (top-10 items are more likely to be purchased).
- Using it in combination with **revenue uplift** ensures that the model not only predicts relevant items but also **maximizes business impact**.

6.4 Outcome:

- By optimizing for NDCG@10, the model ensures that the **most valuable items for each guest appear in the top recommendations**, improving both **guest satisfaction** and **ancillary revenue**