Run Book for Business Metrics and Data Pipelines

Pipeline 1: Profit

- **Primary Owner**: John (Data Analyst)
- **Secondary Owner**: Emily (Data Engineer)
- **Upstream Owners** (Teams): Finance Team
- **Common Issues:**
    - Data discrepancies: Mismatch between financial data sources.
    - Financial data updates: Delayed or incomplete updates.
    - Calculation errors: Errors in profit calculations.
- **Critical Downstream Owners (Teams):** Investor Relations Team
- **SLAs and Agreements**: Data should be updated daily by 9:00 AM. Monthly reports for investors should be delivered by the 5th of the following month.

Pipeline 2: Unit-level Profit for Experiments

- **Primary Owner**: Sarah (Data Scientist)
- **Secondary Owner**: Michael (Data Engineer)
- **Upstream Owners** (Teams): Product Development Team
- **Common Issues**:
    - Experiment design changes: Frequent changes in experiment parameters.
    - Data schema updates: Changes in data structure.
- **Critical Downstream Owners (Teams):** Product Managers
- **SLAs and Agreements**: Data should be updated in real-time. Experiment results should be shared with Product Managers within 24 hours of completion.

Pipeline 3: Aggregate Profit Reported to Investors

- **Primary Owner**: Jennifer (Data Manager)
- **Secondary Owner**: David Clark (Data Engineer)
- **Upstream Owners** (Teams): Profit Pipeline Team
- **Common Issues**:
    - Experiment design changes: Frequent changes in experiment parameters.
    - Data schema updates: Changes in data structure.
- **Critical Downstream Owners (Teams):** Executive Leadership Team
- **SLAs and Agreements:** Monthly reports should be delivered to the Executive Leadership Team by the 5th of the following month.

Pipeline 4: Growth

- **Primary Owner**: Maria Garcia (Data Analyst)
- **Secondary Owner:** Robert Turner (Data Engineer)
- **Upstream Owners (Teams):** Marketing Team
- **Common Issues:**
    - Experiment design changes: Frequent changes in experiment parameters.
    - Data schema updates: Changes in data structure.

- **Critical Downstream Owners (Teams):** Investor Relations Team
- **SLAs and Agreements**: Data should be updated daily by 9:00 AM. Monthly growth reports for investors should be delivered by the 5th of the following month.

Pipeline 5: Daily Growth for Experiments

- **Primary Owner**: Emily Davis (Data Scientist)
- **Secondary Owner**: Daniel White (Data Engineer)
- **Upstream Owners (Teams):** Experimentation Team
- **Common Issues:**
    - Experiment design changes: Frequent changes in experiment parameters.
    - Data schema updates: Changes in data structure.
- **Critical Downstream Owners (Teams):** Product Managers
- **SLAs and Agreements**: Data should be updated in real-time. Experiment results should be shared with Product Managers within 24 hours of completion.

On-Call Schedule

- A fair on-call schedule should be established, considering the following factors:
    - Rotation among primary and secondary owners.
    - Consideration of holidays, weekends, and working hours.
    - Ensuring that each team member has an equal share of on-call responsibilities.

Potential Pipeline Issues

- **Data Quality**: Inaccurate or incomplete data.
- **Schema Changes**: Changes in data structure impacting pipeline processing.
- **Data Pipeline Failures**: Unexpected failures affecting data flow.
- **Experiment Changes:** Frequent changes in experiment parameters.
- **Reporting Errors:** Incorrect data reported to investors.
- **Deadline Misses**: Failure to meet SLAs for data updates and reporting.

Business Requirements

How many total bookings are made on Airbnb?

- Breakdown by geographical regions.
- Determine the source of bookings.

What is the daily rate of new bookings?

- Calculate the percentage of website visitors who convert to making bookings.

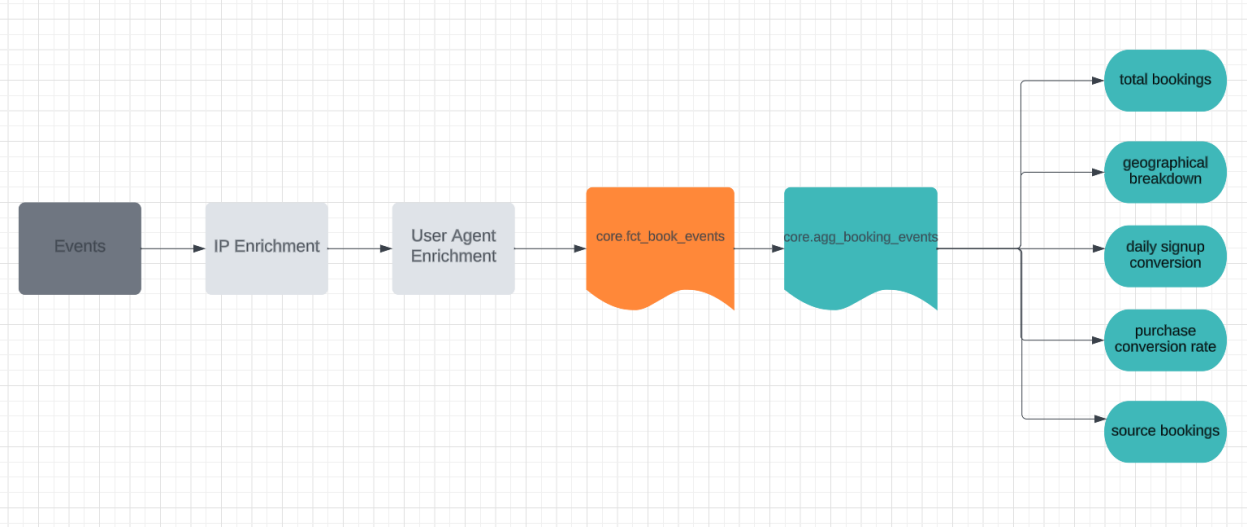
How many users are purchasing Airbnb experiences?

- Calculate the percentage of signups that result in purchases.

Business Metrics

Metric Name	Definition	Is Guardrail?
total_bookings	COUNT(bookings)	YES
geographical_breakdown	COUNT(bookings) GROUP BY region	NO
daily_signup_conversion	COUNT(bookings) / COUNT(website_visitors)	YES
purchase_conversion_rate	COUNT(experiences) / COUNT(signups)	YES
source_of_bookings	COUNT(bookings) GROUP BY referral_source	NO

Flow Diagram



Schemas

core.fct_booking_events: This table contains a list of all booking events on Airbnb, including information about users and bookings.

Column Name	Column Type	Column Comment

user_id	BIGINT	User ID associated with the booking (nullable)
booking_id	BIGINT	Unique identifier for each booking
region	STRING	Geographical region associated with the booking
referral_source	STRING	The source through which the booking was made
booking_timestamp	TIMESTAMP	UTC timestamp for when the booking occurred
other_properties	MAP[string,string]	Additional properties related to the booking
ds	STRING	Date partition for the table

Quality Checks:

- Check for null values in user_id, booking_id, region, referral_source, and booking_timestamp.
- Ensure no duplicate booking IDs.
- Verify that the region values are in the expected format.
- Perform row count checks for different regions and referral sources.
- Enumerate the referral_source values (e.g., direct, search, referral).

Core.agg_booking_events: This table is an aggregated view of all booking events.

Column Name	Column Type	Column Comment
dim_region	STRING	The geographical region
dim_referral_source	STRING	The referral source
m_total_bookings	BIGINT	The total number of bookings for this slice
aggregation_level	STRING	How the aggregation is grouped (e.g., dim_region, dim_referral_source, overall)
ds	STRING	Date partition for the table

Quality Checks:

- Perform row count checks to ensure that the "overall" rollup has the highest count.
- Analyze the seasonality of m_total_bookings.
- Ensure m_total_bookings is greater than a minimum threshold.