Governance Note

Assumptions

This project assumes that historical hotel booking data accurately reflects real world customer behavior patterns and that cancellation behavior remains structurally similar over time. While booking data spans multiple years, it is assumed that key drivers such as lead time, deposit type, and customer history continue to influence cancellation likelihood in comparable ways. Weather data is assumed to be accurate at the daily level and representative of conditions experienced by guests near arrival dates. Importantly, the system assumes that weather acts as contextual information rather than a causal driver of cancellations. Predictions are interpreted as probabilistic risk indicators, not deterministic outcomes. The dashboard and model are designed under the assumption that stakeholders will use results to inform planning and monitoring rather than automated decision making.

Data Ethics

From an ethical standpoint, this project avoids the use of sensitive personal identifiers. The hotel booking dataset does not contain names, addresses, payment details, or other directly identifiable information. Customer attributes such as country, customer type, and repeat guest indicators are used only in aggregate modeling contexts. The machine learning model focuses on behavioral and transactional patterns rather than demographic profiling. Weather features are incorporated to provide environmental context, not to justify or explain individual customer actions. Model outputs are presented as trends and probabilities to reduce the risk of misinterpretation or unfair treatment of specific customer groups. Ethical considerations are reinforced by explicitly communicating that correlation does not imply causation, especially when interpreting weather effects.

Privacy and Security

Data privacy and security are addressed through the use of Google Cloud managed services. All datasets are stored in BigQuery with access restricted via IAM roles at the project and dataset level. Only authorized team members have read or write access to production datasets. Streaming weather data is ingested via a Cloud Function using an API key stored securely in environment variables rather than hard coded credentials. Data transmission occurs over HTTPS, and no personally identifiable information is transmitted or stored during the streaming process. Looker Studio dashboards are shared only with authorized viewers and editors. No external users have access to raw datasets or model artifacts.

Failure Playbook

Several failure scenarios were considered and mitigated during system design. If the weather API becomes unavailable or returns incomplete data, the streaming pipeline will fail gracefully without affecting historical BigQuery tables. In this case, dashboards will continue to display historical

trends while real time updates pause temporarily. If malformed or unexpected weather values are ingested, schema validation and aggregation logic in BigQuery reduce the risk of downstream errors. For model related failures, such as degraded prediction performance or unstable feature importance, the system allows retraining using historical data without disrupting dashboard availability. Logging and monitoring through Google Cloud Logging enables identification of ingestion or scheduling failures. In all cases, the system prioritizes transparency and data integrity over uninterrupted real time updates.