

Informe de evaluación

Propósito

El presente informe presenta los resultados de la validación del modelo analítico y la coherencia de los KPIs del dashboard del proyecto TranStats (Bureau of Transportation Statistics, 2011–2021), desarrollado bajo la metodología CRISP-DM. Su propósito es confirmar que el modelo y las visualizaciones reflejan fielmente los objetivos de negocio definidos en la etapa de Business Understanding.

Hallazgos Clave

El modelo demuestra alta calidad y consistencia. Los KPIs principales: puntualidad, causas de demora y cancelaciones, son coherentes entre el dashboard y las consultas SQL, con una coincidencia superior al 99%. Se detectaron valores atípicos en la comparación entre horarios programados y reales, correspondientes a retrasos operativos extremos, los cuales se consideran fenómenos reales y no errores del cálculo.

Conclusión

El modelo está listo para su uso operativo, respaldado por un proceso ETL confiable y KPIs validados. Se recomienda analizar los valores atípicos identificados y continuar con la automatización de cargas para garantizar actualizaciones continuas.

Calidad y Completitud del Modelo

Descripción de la Fuente de Datos

Los datos provienen del portal oficial del Bureau of Transportation Statistics (BTS), específicamente del dataset TranStats – On-Time Performance (2011–2021).

La información fue integrada en un modelo estrella en SQL Server, compuesto por la tabla de hechos fact_flights y las dimensiones dim_airline, dim_airport, dim_fecha y dim_causa_retraso.

Calidad de los Datos (Input)

Durante el proceso ETL se garantiza la **integridad y consistencia** de los registros.

- Los campos de demora contenían valores nulos en 85% de los casos, reemplazados por 0 al representar vuelos sin retraso.
- Los campos de tiempo con valores fuera del rango 1–2400 fueron eliminados (Aprox 0.9% del total).
- Se comprobó que todas las claves foráneas tienen correspondencia con sus dimensiones, asegurando integridad referencial.

Complejidad (Output):

El modelo cubre el período **2011–2021** e incluye la totalidad de registros válidos después del proceso de limpieza (Aprox. 99% del total original). Esto garantiza cobertura temporal y geográfica completa, permitiendo un análisis preciso de puntualidad, cancelaciones y causas de retraso.

Coherencia entre KPIs del Dashboard y Datos Reales

Método de Validación:

La coherencia de los KPIs fue verificada mediante consultas SQL directas sobre la tabla fact_flights, replicando la lógica de cálculo aplicada en el dashboard.

Revisión por Indicador:

- **% Puntualidad (Éxito):** Los resultados son coherentes. Se observó una ligera variación en algunos aeropuertos por **salidas antes del horario programado**, lo cual no afecta la validez del KPI.
- **Distribución de Retrasos (Éxito):** El KPI refleja correctamente las principales causas de demora, destacando **Delay** y como factores dominantes.
- **Total de Cancelaciones y Desvíos (Éxito):** Los valores coinciden con las consultas de conteo simples (**CANCELLED = 1**, **DIVERTED = 1**).
- **Comparación Programado vs. Real (Atípico):** Se identificaron outliers con demoras de varias horas. No se trata de errores de cálculo, sino de eventos reales que requieren análisis separado para evitar distorsionar los promedios.

Comparación vs. Objetivos del Business Understanding

Objetivo de Negocio 1 – Medir la eficiencia operacional

Validado mediante los KPIs de % Puntualidad y Distribución de Retrasos, que permiten medir el desempeño de aerolíneas y aeropuertos con base en tiempos reales.

Objetivo de Negocio 2 – Identificar oportunidades de mejora

Cumplido mediante el Ranking de Aerolíneas y el análisis de causas de demora, que señalan las áreas donde una intervención operativa podría reducir significativamente los retrasos.

Conclusión sobre el Negocio

Los KPIs seleccionados responden directamente a las preguntas de negocio definidas, evidenciando un alineamiento completo entre el análisis técnico y los objetivos estratégicos del proyecto.

Validación Cruzada por Roles (Cross-Checking)

Roles Involucrados:

- Data Administrator: Verificó la calidad, limpieza y consistencia del modelo.
- Data Engineer: Validó el proceso ETL y la carga en SQL Server.
- BI Analyst: Revisión del dashboard y correspondencia con los datos fuente.
- Data Analyst: Comprobó coherencia estadística y análisis de outliers.

Proceso:

El BI Analyst validó el ranking de aerolíneas y puntualidad, confirmando que los resultados reflejan la realidad operativa (ejemplo: Southwest Airlines Co. como aerolínea con mayor volumen de vuelos).

El Data Engineer y el Data Administrator corroboraron que las métricas del dashboard coinciden con las consultas directas, garantizando trazabilidad completa.

Conclusión y Próximos Pasos

Conclusión:

- El modelo analítico implementado demuestra alta fiabilidad y consistencia, cumpliendo los objetivos de negocio y técnicos del proyecto.
- Los KPIs validados reflejan adecuadamente la realidad operacional del transporte aéreo entre 2011 y 2021.

Mejoras Futuras:

La Lista de Mejoras Futuras fue elaborada con base en los hallazgos del KPI Comparación Programado vs. Real y otros puntos menores detectados.

Se recomienda:

- Implementar filtros para outliers en los KPIs de retraso.
- Automatizar el proceso ETL para reducir tiempo de actualización.
- Incluir datos meteorológicos externos para fortalecer el análisis causal.