

UK Train Ticket Data Analysis Report

UK Train Ticket Data Analysis Report

أولاً: فهم البيانات (Data Understanding)

1. نظرة عامة على الـ Dataset

• عدد الصفوف: 31,653

• عدد الأعمدة: 18

• (لا توجد قيم مفقودة) Nulls = 0

• تحتوي على بيانات عمليات شراء تذاكر قطارات في المملكة المتحدة، وتشمل معلومات عن الرحلات، الأسعار، طرق الدفع، حالة الرحلات، والاستردادات.

2. الأعمدة ومعانيها التفصيلية

العمود	الوصف	ملاحظات تحليلية	علاقته بالأعمدة الأخرى
Transaction ID	رقم العملية الفريد	Primary Key	يستخدم لربط كل عملية شراء ببقية التفاصيل
Date of Purchase	تاريخ شراء التذكرة	"Date of Journey"	لتحليل سلوك الشراء عبر الزمن يرتبط بـ "Date of Journey" لتحديد الفاصل بين الحجز والسفر
Time of Purchase	توقيت الشراء	"Date of Purchase"	لفهم أوقات الذروة في الشراء يمكن ربطه بـ "Date of Purchase"
Purchase Type (person)	نوع الشراء	Payment Method	لمعرفة القناة الأكثر استخداماً يؤثر على
Payment Method	طريقة الدفع	Customer Demographic	يتأثر بنوع الشراء
Railcard	نوع البطاقة المخفضة	Price	يؤثر مباشرة على الإيرادات
Ticket Class	فئة التذكرة	First / Standard	يتفاعل مع Price و Demand
NA	غير ملحوظ	مؤشر على مستوى الخدمة	مؤشر على مستوى الخدمة

Price و Journey يتفاعل مع	لفهم أنماط السفر	نوع التذكرة (/) Single / Return (Season)	Ticket Type
يعتمد على الفئة، المسافة، نوع التذكرة	متغير رقمي أساسي للتحليل المالي	سعر التذكرة	Price
Arrival Destination يمكن ربطها بـ	جغرافيًّا مهمة لتحليل الشبكة	محطة الانطلاق	Departure Station
مرتبط بـ Departure Station	تحدد الوجهة؛ تساعد على معرفة المسارات الأعلى استخدامًا	محطة الوصول	Arrival Destination
- محور زمني للتحليل؛ يُستخدم لتوقع الطلب اليومي	تاريخ الرحلة	Date of Journey	
Arrival Time يُقارن بـ	لتحليل الزمني	موعد المغادرة المجدول	Departure Time
Actual Arrival Time مقارنة بـ	يساهم في حساب مدة الرحلة	موعد الوصول المجدول	Arrival Time
Arrival Time الفرق بين Actual و يعطي التأخير	يستخدم في قياس التأخير	وقت الوصول الفعلي	Actual Arrival Time
Refund Request يؤثر على	يقيس كفاءة التشغيل	حالة الرحلة (/) On time (Delayed / Cancelled)	Journey Status
Journey Status يُستخدم مع	مهم لتحليل الأداء التشغيلي	سبب التأخير (/) Weather (Technical / etc)	Reason for Delay
Price و Journey Status يتأثر بـ	لقياس رضا العملاء	هل تم طلب استرداد؟ (/) Yes (No)	Refund Request

ثانياً: فهم بيئة العمل (Business Context)

محاور رئيسية: 3

قطاع السكك الحديدية البريطاني يعتمد على Revenue (الإيرادات) – بيع التذاكر ومعدل استخدام الفئات

الاعلى سعراً.

• Customer Satisfaction (رضاء العملاء) – يعتمد على دقة المواعيد وجودة الخدمة.

• Operational Efficiency (التشغيل) – تأخير القطارات واسترداد الأموال.

أهداف المشروع:

• زيادة الإيرادات من خلال فهم أنماط الشراء والسفر.

• تحسين دقة التنبؤ بالطلب اليومي على الرحلات والتسعير.

• تحليل الأداء لكل محطة وخط سير.

• تقليل نسبة التأخير واسترداد الأموال.

• بناء نماذج تنبؤية للطلب والإيرادات.

مؤشرات الأداء الرئيسية (KPIs)

Operational KPIs:

- On-Time Performance (%)
- Average Delay (Minutes)
- Journey Completion Rate (%)
- Cancellation Rate (%)
- Delay by Route / Station
- Turnaround Time
- Operational Efficiency Score

Customer & Ticketing KPIs:

- Total Tickets Sold
- Passenger Volume per Route / Class
- Refund Request Rate
- Average Ticket Price
- Passenger Segmentation (by Class / Type / Railcard)

Financial KPIs:

- Total Revenue
- Revenue by Route / Station / Class
- Revenue per Passenger (Yield)
- Refund Cost %
- Fare Distribution
- Revenue Growth (%)

Service Quality KPIs:

- Customer Complaints Count
- Refund Reasons Analysis
- On-Time Reliability Trend
- Journey Rating (if available)

Forecasting KPIs:

- Passenger Demand Forecast
- Revenue Forecast
- Seasonal Demand Patterns
- Route Demand Prediction

ثالثاً: العلاقة بين الأعمدة (Data Relationships)

- السعر يتغير حسب الفئة، نوع التذكرة، نوع بطاقة التخفيض: $\text{Price} \leftrightarrow \text{Ticket Class} / \text{Ticket Type} / \text{Railcard}$.
- الفرق بين الوقتين يحدد حالة الرحلة: $\text{Journey Status} \leftrightarrow \text{Actual Arrival Time} - \text{Arrival Time}$.
- كلما زاد التأخير والسعر، زادت احتمالية طلب الاسترداد: $\text{Refund Request} \leftrightarrow \text{Journey Status} / \text{Price}$.
- عمليات تستخدم بطاقات أكثر من النقد: $\text{Purchase Type} \leftrightarrow \text{Payment Method: Online}$.
- لتحديد أكثر المسارات شيوعاً: $\text{Departure Station} \leftrightarrow \text{Arrival Destination}$.

رابعاً: خطة تنظيف ومعالجة البيانات (Data Cleaning & Preprocessing)

- تحويل الأعمدة الزمنية إلى datetime: Date of Purchase, Date of Journey, Departure Time.

.Arrival Time, Actual Arrival Time

- إنشاء عمود Delay Duration = Actual Arrival – Arrival Time
- استخراج Month, Weekday, Hour (Arrival Departure) لتحليل الأنماط الزمنية.
- التحقق من القيم غير المنطقية (Negative prices).
- توحيد القيم النصية (مثلاً: Online / online / On-line).
- تحديد الأعمدة المشتقة (مثل Revenue per Class).

١. سلوك العملاء (Customer Behavior Analysis)

الأدوات: Python (pandas, matplotlib) / SQL

أسئلة تحليلية:

• أكثر ساعات الشراء.

- الفرق بين أيام العمل وعطلة نهاية الأسبوع.
- أكثر أيام الأسبوع للشراء.
- القناة الأكثر استخداماً ومعدل رضا (In-person وOnline و اختلافها بين).
- العلاقة بين وقت الشراء ووقت الرحلة.
- اختلاف سلوك الشراء بين العملاء مقابل Online الوجهات الأكثر شراء المدن الكبرى On-site.
- متوسط عدد التذاكر لكل عميل في الشهر.
- العلاقة بين نوع البطاقة وفئة السعر.
- تفضيلات بطاقات التخفيض.
- ارتباط طريقة الدفع بطلبات الاسترداد.
- ما أكثر محطات المغادرة والوصول شيوعاً؟
- ما متوسط سعر التذكرة لكل نوع (Advance, Off-Peak, Anytime)؟
- ما نسبة التذاكر المباعة أونلاين مقابل في المحطة؟
- ما أكثر وسيلة دفع استخداماً؟
- ما الأيام أو الأشهر ذات أعلى حجوزات؟
- ما العلاقة بين وقت الشراء ووقت الرحلة؟
- هل العملاء يفضلون الحجز Online أم في المحطة؟
- ما أنواع التذاكر الأكثر استخداماً؟
- ما فئة التذاكر (First / Standard) التي تحقق رضا أعلى؟
- هل حاملي Railcards يشترون أكثر Online؟
- هل الحجز المبكر يقلل احتمال التأخير؟
- هل الطلب يزيد في مواسم أو عطلات معينة؟

2. الأداء التشغيلي (Operational Efficiency Analysis)

الأدوات: SQL / Python

أسئلة تحليلية:

- متوسط مدة التأخير بالدقائق.

• نسبة الرحلات في موعدها مقابل المتأخرة أو الملغاة.

• الأسباب الأكثر شيوعاً للتأخير.

• اختلاف معدلات التأخير بين أيام الأسبوع والعطل.

• الفترات الزمنية الأعلى تأثيراً.

• علاقة نوع المحطة بدرجة التأخير.

• نسبة الرحلات التي تجاوزت 15 دقيقة.

• مقارنة انتظام الرحلات بين فئات التذاكر.

• خطوط سير أو شركات تشغيل ذات تأخير أعلى.

• التأخير وطلبات الاسترداد.

• الطقس وعدد التأخيرات المفقود سنوياً.

• ارتباط التأخيرات بمحطات أو أسباب محددة.

• ما نسبة الرحلات On Time / Delayed / Cancelled؟

• ما متوسط التأخير بالدقائق؟

• ما أكثر الأسباب شيوعاً للتأخير؟

• هل التأخيرات تحدث أكثر في أيام معينة أو مواسم محددة؟

• هل الأداء يختلف حسب شركة التشغيل أو المحطة؟

• هل الرحلات الطويلة تتأخر أكثر من القصيرة؟

• ما العلاقة بين درجة التذكرة واحتمال التأخير؟

• ما نسبة الركاب الذين طلبوا Refund؟

• هل ركاب الدرجة الأولى يطلبون استرداداً أكثر؟

• هل سعر التذكرة يؤثر على احتمالية طلب Refund؟

• هل طلبات Refund تزيد في مواسم معينة؟

• هل الرحلات المتأخرة أكثر عرضة لطلبات استرداد؟

• ما الأسباب الأكثر شيوعاً لطلبات Refund؟

3. الإيرادات والتحليل المالي (Revenue & Sales Analysis)

أسئلة تحليلية:

• فئة التذاكر الأكثر إيراداً.

• المحطات الأعلى إيراداً.

• إجمالي الإيرادات اليومي، الأسبوعي، الشهري.

• نوع التذكرة الأكثر ربحاً.

• تأثير بطاقات الخصم على الإيرادات.

• الإيرادات المتأثرة بطلبات الاسترداد.

• متوسط الإيراد لكل رحلة.

• متوسط سعر التذكرة لكل فئة.

• العلاقة بين الأسعار والمبيعات.

• الاتجاه العام للإيرادات (Seasonality).

• تأثير المواسم على الإيرادات.

• المحطات الأعلى دخلاً لكل رحلة.

• تأثير التأخيرات على الإيرادات.

• ما إجمالي الإيرادات الشهرية / اليومية؟

• ما المحطات أو الخطوط الأعلى ربحية؟

• هل التأخيرات تؤثر على الإيرادات؟

• ما فئة التذاكر الأعلى ربحاً؟

• هل بطاقات الخصم تقلل الإيرادات الكلية؟

• ما العلاقة بين سعر التذكرة ومتوسط مدة التأخير؟

• هل المواسم (صيف/شتاء) تؤثر على المبيعات؟

4. التنبؤات (Forecasting & Demand Planning)

الأدوات: Python (scikit-learn, statsmodels)

أسئلة تحليلية:

- توزيع الطلب بين فئات التذاكر.
- الإيراد المتوقع لكل يوم.
- توقع عدد الرحلات في الشهر القادم.
- تأثير زيادة التأخير على الإيرادات.
- تأثير الأعياد والعطل.
- النمط الأسبوعي للرحلات.
- الاتجاه الموسمى في عدد الرحلات.
- تقدير الإيرادات المحتملة عند تحسين الالتزام بالمواعيد.
- تأثير العروض والتخفيضات.
- المدن الأعلى نمواً في عدد الركاب.
- التنبؤ بطلب الاسترداد المستقبلي.
- توقع عدد الرحلات في الشهر القادم.
- توقع الإيراد لكل يوم في الشهر القادم.
- توقع الطلب على فئات التذاكر المختلفة.
- تحليل التأثير الموسمى على الطلب والإيرادات.
- توقع حجم الطلب في خطوط معينة (Route Demand).
- تحديد أكثر الأسباب التي تستحق الاستثمار في حلها.
- بناء نموذج تنبؤ للرحلات المتوقع تأخيرها (Predictive Delay Model).
- تقييم أثر تحسين الالتزام بالمواعيد على الإيرادات.

5. رؤى تحليلية إضافية (Cross-sectional / Combined Insights)

- Origin-Destination Matrix: خريطة تفاعلية لكتافة الرحلات بين المحطات.

- مقارنة القتوات في الإيرادات والرضا .
- منحنى زمني يجمع بين الإيرادات و عدد الرحلات يومياً.
- علاقة تأخيرات الرحلات بمستوى المبيعات .
- Performance Score لكل محطة (Punctuality + Revenue + Refund Rate).

الوصيات التحليلية (Strategic Recommendations)

- تحسين أداء المحطات الأكثر تأخيراً أولاً.
- تقليل أعطال Signal Failure بنسبة 20 % لتحسين الأداء العام.
- تطبيق تعويض تلقائي لتقليل طلبات Refund اليدوية.
- استهداف فئات التذاكر الأعلى ربحاً بتسعير مرن.
- التركيز على أشهر الذروة لتحسين الإيرادات التشغيلية.
- تطوير Dashboard يربط بين الأداء المالي والتشغيل والرضا.