

تحليل بيانات مشاركة الدراجات في المدينة

مشروع تخرج في علوم البيانات

مقدمة المشروع

التحدي

في عصر النقل المستدام، أصبحت أنظمة مشاركة الدراجات جزءاً أساسياً من البنية التحتية الحضرية. يواجه هذا المشروع تحدي تحليل كميات كبيرة من بيانات رحلات الدراجات لفهم أنماط الاستخدام وتأثير العوامل البيئية على الربحية.

الهدف

يهدف هذا المشروع إلى: - فهم الأنماط الزمنية والمكانية لاستخدام الدراجات - تحليل تأثير الطقس على تكلفة الرحلات والإيرادات - تحديد المناطق عالية الاستخدام والأنماط المتكررة - تطوير نماذج تنبؤية للإيرادات المستقبلية - اكتشاف مجموعات المستخدمين بناءً على سلوكهم

مراحل العمل

1. تحميل وتنظيف البيانات (Block 1-2)
2. دمج ملفات البيانات المتعددة
3. معالجة القيم المفقودة والبيانات الشاذة
4. توحيد التنسيقات والأنواع
5. هندسة الميزات (Block 3-4)
6. إنشاء متغيرات زمنية (اليوم، الشهر، الموسم)
7. حساب المسافات والسرعات
8. إضافة ميزات الطقس والموقع
9. التحليل الاستكشافي (Block 5)
10. تحليل التوزيعات والإحصائيات الوصفية
11. استكشاف العلاقات بين المتغيرات
12. إنشاء خرائط حرارية للمناطق
13. التنبؤ الزمني (Block 6)
14. استخدام نموذج Prophet للتنبؤ بالإيرادات
15. تحليل المكونات الموسمية والاتجاهات
16. اكتشاف الأنماط (Block 7)
17. تطبيق خوارزميات التجميع (K-Means, DBSCAN, Agglomerative)
18. تحليل خصائص كل مجموعة

الأدوات المستخدمة

- Python: اللغة الأساسية للتحليل
- Pandas: معالجة البيانات والتحليل

- **GeoPandas**: التحليل المكاني
- **Prophet**: التنبؤ الزمني
- **Plotly**: الرسوم البيانية التفاعلية
- **Scikit-learn**: خوارزميات التعلم الآلي
- **NumPy**: العمليات الحسابية

الرؤى الرئيسية

التوزيع الزمني والمكاني للرحلات

أفضل محطات البداية

- **الأنماط الزمنية**: ذروة الاستخدام في ساعات الذروة (8-9 صباحاً، 5-6 مساءً)
- **التوزيع الأسبوعي**: زيادة الاستخدام في أيام العمل مقارنة بالعطلات
- **التوزيع الموسمي**: ارتفاع الاستخدام في الربيع والصيف

العلاقة بين التكلفة والمدة والطقس

توزيع تكلفة الرحلات

- **التكلفة والمدة**: علاقة إيجابية قوية (معامل الارتباط: 0.85)
- **تأثير الطقس**: انخفاض التكلفة في الأيام الممطرة والباردة
- **السرعة المتوسطة**: 12-15 كم/ساعة للرحلات العادية

الأنماط المتكررة والمناطق عالية الاستخدام

خريطة حرارية للمناطق

- **المناطق المركزية**: أعلى كثافة استخدام
- **المناطق السكنية**: أنماط منتظمة صباحاً ومساءً
- **المناطق التجارية**: استخدام متواصل طوال اليوم

سلوك المستخدمين حسب العضوية ونوع الدراجة

تحليل العضوية ونوع الدراجة

- الأعضاء السنويون: رحلات أطول وأكثر تكراراً
- المستخدمون العاديون: رحلات قصيرة ومتفرقة
- الدراجات الكهربائية: شعبية متزايدة، خاصة للمسافات الطويلة

تأثير القرب من المناطق المركزية ووسائل النقل

تحليل المناطق المركزية

- القرب من المترو: زيادة بنسبة 40% في الاستخدام
- المناطق المركزية: رحلات أقصر وأكثر تكراراً
- الضواحي: رحلات أطول وأقل تكراراً

نتائج التعلم الآلي

التنبؤ بالإيرادات باستخدام Prophet

تنبؤ الإيرادات

النتائج الرئيسية: - نمو سنوي متوقع: 15-20% - أنماط موسمية واضحة (ذروة في الصيف) - تأثير الأحداث الخاصة والعطلات

مكونات التنبؤ

تحليل المكونات: - الاتجاه العام: نمو مستمر - الموسمية الأسبوعية: ذروة في أيام العمل - الموسمية السنوية: ذروة في يونيو-أغسطس

تحليل المجموعات (Clusters)

مجموعة K-Means

تحليل المجموعات - K-Means

المجموعات المكتشفة: 1. المستخدمون المنتظمون: رحلات قصيرة ومتكررة
2. المسافرون لمسافات طويلة : رحلات طويلة ونادرة 3. المستخدمون
المتوسطون: رحلات متوازنة

مجموعة DBSCAN

تحليل المجموعات - DBSCAN

الأنماط المكتشفة: - المجموعة الأساسية : 85% من المستخدمين -
المجموعات الفرعية: أنماط متخصصة (الطلاب، الموظفون، السياح)

التحليل الهرمي

الشجرة الهرمية

التقسيم الهرمي: - 3 مستويات رئيسية من التجميع - تمييز واضح بين الأنماط
المختلفة

التحديات والحلول

مشاكل الأداء (حجم البيانات)

التحدي: - حجم البيانات الأصلي: +2 مليون رحلة - وقت المعالجة: ساعات لكل
تحليل - استهلاك الذاكرة: +8 جيجابايت

الحل: - استخدام عينات مصغرة (10,000 رحلة) - تحسين خوارزميات المعالجة -
استخدام تنسيق Parquet للضغط

اختبارات التحقق واستخدام أدوات فعالة

الاستراتيجية: - اختبارات شاملة لكل مرحلة - التحقق من جودة البيانات -
مقارنة النتائج مع البيانات الكاملة

الأدوات المستخدمة: - Pandas للتحليل السريع - Plotly للرسوم التفاعلية -
Scikit-learn للتعلم الآلي

الخلاصة والتوصيات

أهم النتائج

1. الأنماط الزمنية: ذروة واضحة في ساعات الذروة
2. التأثير المكاني: تركيز في المناطق المركزية
3. تأثير الطقس: انخفاض الاستخدام في الظروف السيئة
4. التنبؤ: نمو مستمر بنسبة 15-20% سنوياً
5. المجموعات: 3 أنماط رئيسية للمستخدمين

التوصيات المستقبلية

قصيرة المدى (3-6 أشهر)

- تحسين التوزيع: زيادة الدراجات في المناطق عالية الاستخدام
- التسعير الديناميكي: تعديل الأسعار حسب الطلب والطقس
- التسويق المستهدف: حملات مخصصة لكل مجموعة مستخدمين

متوسطة المدى (6-12 شهر)

- توسيع الشبكة: إضافة محطات في المناطق الجديدة
- تحسين التطبيق: ميزات تنبؤية للمستخدمين
- الشراكات: تعاون مع وسائل النقل العام

طويلة المدى (+1 سنة)

- الذكاء الاصطناعي: نماذج تنبؤية متقدمة
- التحليل في الوقت الفعلي: مراقبة مستمرة للأنماط
- التوسع الجغرافي: الوصول لمدن جديدة

الأفكار القابلة للتنفيذ

1. نظام التنبؤ الذكي

2.	تنبؤ بالطلب في كل محطة
3.	إعادة توزيع تلقائية للدراجات
4.	برنامج الولاء المتقدم
5.	مكافآت مخصصة لكل مجموعة
6.	خصومات ديناميكية
7.	تحليل التنقل المتكامل
8.	دمج بيانات وسائل النقل الأخرى
9.	تحليل رحلات الباب إلى الباب
10.	الاستدامة البيئية
11.	تتبع الانبعاثات المحفوظة
12.	تقارير التأثير البيئي

ملخص الإحصائيات الرئيسية

القيمة	المقياس
10,000	إجمالي الرحلات
18 دقيقة	متوسط مدة الرحلة
3.50 ريال	متوسط التكلفة
2.8 كم	متوسط المسافة
65%	نسبة الأعضاء السنويين
35%	نسبة الدراجات الكهربائية

هيكل المشروع

```
    /final_project
      README.md —|
      report_arabic.md —|
      /figures/final_report —|
        top_start_stations.png_01 —| |
        member_bike_sunburst.png_02 —| |
        trip_duration_histogram.png_06 —| |
        trip_cost_distribution.png_10 —| |
        daily_revenue_trend.png_11 —| |
        geohash_heatmap.png_13 —| |
        cbd_usage_analysis.png_17 —| |
        revenue_by_weather.png_19 —| |
        forecast_plot.png —| |
        forecast_components.png —| |
        kmeans_pca.png —| |
        dbscan_tsne.png —| |
        agglomerative_dendrogram.png —L |
      /engineered_data —|
        spatial_features_block4.parquet —| |
        eda_block5_outputs.parquet —| |
        forecast_block6.csv —| |
        clustered_block7.parquet —L |
      /test_scripts —L
        test_block4.py —|
        test_block5.py —|
        test_block6.py —|
        test_block7.py —L
```

تم إنجاز هذا المشروع كجزء من متطلبات تخرج علوم البيانات - 2024