



الجمهورية العربية السورية

جامعة دمشق

كلية الهندسة المعلوماتية

قسم الذكاء الصناعي

Daily bike trips in WDC

الرحلات اليومية للدراجات في واشنطن

اعداد:

اياد الملك

نوري السعدون

حمزة محارب

19/6/2025

23/12/1446

ملخص المشروع :

يهدف هذا المشروع إلى تحليل أنماط استخدام الدراجات اليومية ضمن نظام مشاركة الدراجات في العاصمة واشنطن، وذلك من خلال تقنيات التنقيب في البيانات، تم استخدام عدة مجموعات بيانات تشمل سجلات الرحلات اليومية، ومعلومات الطقس، ومواقع محطات الدراجات، بالإضافة إلى مناطق جغرافية مثل المنطقة التجارية المركزية ومواقف الحافلات ومناطق الوقوف.

يتضمن المشروع مراحل متعددة تبدأ بتنظيف البيانات وتهيئتها، مروراً باستخراج السمات الهندسية مثل مدة الرحلة وتكلفتها والمسافة من المناطق الحيوية، ثم تحليل البيانات استكشافياً من خلال الرسوم البيانية والمخططات الإحصائية، بالإضافة إلى تحليل مكاني وزمني، و انتهاءً بالقيام بعمليات اصياد و اكتشاف الانماط.

كما يتضمن المشروع نمذجة للسلاسل الزمنية باستخدام Prophet لتوقع الإيرادات المستقبلية، وتطبيق خوارزميات التجميع (Clustering) لاكتشاف أنماط الاستخدام المختلفة، تهدف هذه التحليلات إلى دعم اتخاذ القرار من خلال توفير رؤى واضحة حول سلوك المستخدمين وتأثير العوامل البيئية والمكانية والزمنية على استخدام الدراجات.

1. المقدمة :

1.1. خلفية المشكلة :

شهدت السنوات الأخيرة زيادة ملحوظة في استخدام وسائل التنقل النظيفة في المدن الكبرى، ومن أبرزها أنظمة مشاركة الدراجات التي تتيح للمستخدمين استئجار دراجات للتنقل داخل المدينة، تُعد العاصمة الأمريكية واشنطن واحدة من المدن الرائدة في هذا المجال، حيث تعتمد على نظام مشاركة دراجات يغطي مختلف المناطق ويُستخدم من قبل سكان المدينة والزوار على حدّ سواء.

ومع تنامي حجم البيانات الناتجة عن هذا النظام يومياً (مثل بيانات الرحلات، الطقس، المحطات الجغرافية)، أصبح تحليل هذه البيانات ضرورة لفهم سلوك المستخدمين وتحسين الخدمات المقدّمة.

1.2. أهمية المشروع :

يساهم هذا المشروع في دعم اتخاذ القرار فيما يتعلّق بتخطيط التنقل الحضري، من خلال استخراج رؤى عملية مبنية على تحليل فعال للبيانات، فهم تأثير العوامل مثل الطقس، الموقع، نوع المستخدم، ونوع الدراجة على أنماط الاستخدام، يُمكن الجهات المختصة من تحسين توزيع المحطات، تسعير الخدمة، والتخطيط للتوسّع في المناطق ذات الطلب المرتفع.

1.3. التحديات في المشروع :

يتضمن المشروع عدّة تحديات تقنية وتحليلية، من أبرزها:

- دمج مجموعات بيانات متعدّدة من مصادر مختلفة مع الحفاظ على التناسق والدقة.
- معالجة البيانات الجغرافية، بما في ذلك التحويل بين أنظمة الإحداثيات المختلفة.
- حساب تكلفة الرحلات بدقة بناءً على شروط متعددة ومعقدة تعتمد على نوع الدراجة ومدة الاستخدام والمنطقة.

- **نمذجة التنبؤ** باستخدام سلاسل زمنية، وهو ما يتطلب اختيار النموذج المناسب وضبطه.

1.4. الاهداف و النتائج المرجوة من المشروع :

يسعى المشروع إلى تحقيق مجموعة من الأهداف التحليلية والفنية، وهي:

- **انشاء مجموعة بيانات** من خلال الدمج بين عدة مجموعات صغيرة و معالجتها .
- **هندسة السمات** من خلال:
 - حساب مدة الرحلة
 - حساب التكلفة الحقيقية لكل رحلة وفقاً لشروط محددة
 - استخراج خصائص مكانية (مثل القرب من المنطقة التجارية أو مواقف المترو)
- **تحليل البيانات استكشافياً** عبر الرسوم البيانية لتحديد أنماط الاستخدام حسب الوقت، النوع، الموقع، والطقس.
- **تحليل مكاني** لتحديد المناطق ذات النشاط المرتفع والمنخفض باستخدام الخرائط الحرارية والتهشير الجغرافي.
- **تحليل تأثير الطقس** على سلوك المستخدمين والإيرادات.
- **نمذجة السلاسل الزمنية** لتوقع الإيرادات المستقبلية.
- **تطبيق خوارزميات التجميع (Clustering)** لاكتشاف أنماط خفية في الاستخدام.

2. وصف مجموعات البيانات :

2.1. رحلات الدراجات اليومية (Daily Bike Trips) :

- **الوصف** : تحتوي هذه المجموعة على تفاصيل جميع الرحلات اليومية التي تم تنفيذها بواسطة نظام مشاركة الدراجات.
- **حجم البيانات** : عدد الصفوف: 6114323 – عدد الأعمدة: 13
- **تمثيل البيانات** : تشمل البيانات عدة عناصر مهمة مثل وقت الانطلاق والانتهاء، المحطة، نوع الدراجة، ونوع المستخدم

2.2. معلومات الطقس اليومية (Daily weather info) :

- **الوصف** : تتضمن هذه المجموعة معلومات الطقس لكل يوم في واشنطن العاصمة تساهم هذه البيانات في تحليل تأثير الطقس على أنماط استخدام الدراجات اليومية.
- **حجم البيانات** : عدد الصفوف: 366 – عدد الأعمدة: 17
- **تمثيل البيانات** : تشمل عناصر مثل درجة الحرارة، سرعة الرياح، نسبة الرطوبة، وحالة السماء .

2.3. معلومات المحطات (Stations info) :

- **الوصف** : تحتوي هذه المجموعة على معلومات تفصيلية عن محطات الدراجات، تُستخدم هذه البيانات لربط الرحلات بالمواقع وتحليل النشاط المكاني ضمن المدينة.
- **حجم البيانات** : عدد الصفوف: 794 – عدد الأعمدة: 29
- **تمثيل البيانات** : تشمل عناصر مثل اسم المحطة، الموقع الجغرافي (إحداثيات)، وسعة المحطة.

2.4. جدول المحطات (Stations table) :

- **الوصف :** يمثل هذا الجدول نسخة منظمة من معلومات المحطات مع معرفات تسلسلية، تُستخدم لربطها مع بيانات الرحلات بدقة أكبر
- **حجم البيانات :** عدد الصفوف: 821 – عدد الأعمدة: 2
- **تمثيل البيانات :** يحوي اسم المحطة و المعرف الخاص بها

2.5. مضلع المنطقة التجارية المركزية (CBD Polygon) :

- **الوصف :** تُستخدم هذه البيانات في التحليل المكاني لتحديد مدى تركّز النشاط داخل قلب المدينة ، ، وتُستخدم لتحليل ما إذا كانت الرحلة تبدأ أو تنتهي داخل المنطقة الحيوية.
- **حجم البيانات :** عدد الصفوف: 1 – عدد الأعمدة: 11
- **تمثيل البيانات :** هذه المجموعة تمثل الحدود الجغرافية للمنطقة التجارية المركزية في واشنطن العاصمة.

2.6. محطات حافلات المترو (Metro Bus Stops) :

- **الوصف :** تتضمن هذه المجموعة معلومات الطقس لكل يوم في واشنطن العاصمة تساهم هذه البيانات في تحليل تأثير الطقس على أنماط استخدام الدراجات اليومية.
- **حجم البيانات :** عدد الصفوف: 10044 – عدد الأعمدة: 79
- **تمثيل البيانات :** تتضمن هذه المجموعة مواقع محطات حافلات المترو ضمن المدينة، تشمل إحداثيات كل محطة.

2.7. محطات الحافلات المكوكية (Shuttle Bus Stops) :

- **الوصف :** تُساعد في تحديد علاقة الرحلات بخطوط النقل المخصصة غير العامة، وتُستخدم في التحليل المكاني للرحلات.
- **حجم البيانات :** عدد الصفوف: 102 – عدد الأعمدة: 29
- **تمثيل البيانات :** تتضمن هذه البيانات مواقع محطات الحافلات الخاصة

2.8. مناطق الوقوف السكني والزوار (Residential and Visitor Parking Zones) :

- **الوصف :** تُستخدم لربط محطات الدراجات بالمناطق السكنية لتحديد مدى استخدامها من قبل السكان أو الزوار ، وتشمل تقسيماً جغرافياً تفصيلياً.
- **حجم البيانات :** عدد الصفوف: 40 – عدد الأعمدة: 14
- **تمثيل البيانات :** تمثل هذه البيانات المناطق المخصصة للوقوف السكني والزوار ضمن المدينة

3. تنظيف البيانات والمعالجة الأولية (Data Cleaning and Preprocessing) :

تضمنت هذه المرحلة عدة خطوات لمعالجة القيم المفقودة، تنسيق أنواع البيانات، ودمج مجموعات البيانات المختلفة.

3.1. معالجة القيم المفقودة في بيانات رحلات الدراجات اليومية :

المشكلة : لوحظ وجود عدد كبير من القيم المفقودة في أعمدة أساسية ضمن مجموعة بيانات رحلات الدراجات اليومية (Daily Bike Trips) ، تحديداً في:

- start_station_name
- start_station_id

- end_station_name
- end_station_id

هذه القيم المفقودة كانت ستحول دون إجراء تحليل دقيق لأنماط الاستخدام حسب المحطة، وحسابات المسافة، والربط بالمناطق الجغرافية.

الحل المتبع: لمعالجة هذه المشكلة، تم استخدام **الوصلة المكانية المتقاربة (Spatial Join Nearest)** لملء القيم المفقودة، تم تحويل بيانات الرحلات وبيانات المحطات إلى اطرادات بيانات جغرافية (GeoDataFrames) باستخدام الإحداثيات الجغرافية (خط الطول وخط العرض)، بعد ذلك تم تحديد أقرب محطة لكل نقطة انطلاق ووصول مفقودة.

3.2. معالجة البيانات المكررة والقيم المفقودة في مجموعات البيانات الأخرى :

- **إزالة الصفوف المكررة:** تم التحقق من كل مجموعة بيانات وإزالة أي صفوف مكررة لضمان عدم تكرار المعلومات.
- **فحص القيم المفقودة:** تم إجراء فحص شامل للقيم المفقودة في جميع الأعمدة، لحسن الحظ، لم تحتوي هذه المجموعات على قيم مفقودة ذات أهمية أو تأثير كبير على التحليل، وبالتالي لم تتطلب معالجة إضافية .

3.3. توحيد أنواع البيانات الزمنية :

لضمان التناسق والدقة في التحليلات الزمنية وربط البيانات، تم تحويل أعمدة التاريخ والوقت في مجموعات البيانات الرئيسية إلى تنسيق Datetime

- **مجموعة بيانات الطقس (weather_df):** تم تحويل عمود datetime إلى نوع بيانات Datetime وتسميته date، هذا يضمن إمكانية التعامل مع التواريخ بسلاسة لإجراء عمليات الدمج والتحليل الزمني.
- **مجموعة بيانات الرحلات (trips_df):** تم تحويل عمود started_at إلى start_time وعمود ended_at إلى end_time كلاهما بتنسيق Datetime هذا يسهل عمليات حساب مدة الرحلة والتحليلات الزمنية الأخرى.

3.4. التأكد من أنظمة الإحداثيات المرجعية (CRS) وربط البيانات:

لضمان دقة التحليلات المكانية، تم التأكد من أن جميع طبقات البيانات الجغرافية تستخدم نظام الإحداثيات المرجعية الموحد EPSG:4326 (CRS) ، إذا لم يكن أي من GeoDataFrames بهذا النظام، فقد تم تحويله إليه.

كما تم إجراء عمليات دمج حاسمة بين مجموعات البيانات المختلفة:

- **ربط المحطات بمناطق الوقوف:** تم دمج مجموعة بيانات المحطات (stations_gdf) مع مناطق الوقوف السكنية والزوار (parking_zones_gdf) باستخدام وصلة مكانية (sjoin) لتحديد ما إذا كانت المحطة تقع ضمن منطقة وقوف معينة، هذا أضاف بعداً مكانياً للتحليل حول نوع المناطق التي تقع فيها المحطات.
- **دمج بيانات الطقس مع بيانات الرحلات:** تم استخراج التاريخ فقط من start_time في trips_df ومن date في weather_df، ثم تم دمج مجموعتي البيانات weather_df و trips_df باستخدام التاريخ كعمود ربط مشترك، هذا الدمج أتاح تحليل تأثير الظروف الجوية المختلفة على أنماط استخدام الدراجات.

3.5. معالجة الرحلات خارج حدود واشنطن العاصمة :

المشكلة: لوحظ وجود بعض سجلات الرحلات التي تقع نقاط انطلاقها أو وصولها خارج الحدود الجغرافية المعروفة لواشنطن العاصمة، هذه النقاط تُعتبر قيماً شاذة (Outliers) وقد تؤثر على دقة التحليلات المكانية وتؤدي إلى استنتاجات خاطئة حول أنماط الاستخدام داخل المدينة.

الحل المتبع: تم تحديد الحدود الجغرافية لمدينة واشنطن العاصمة باستخدام "طريقة الصندوق المحيط (Bounding Box Method)"، والتي تحدد الحد الأدنى والأقصى لخطوط الطول والعرض للمدينة على النحو التالي:

- خط العرض الأدنى (DC_LAT_MIN): 38.7916
- خط العرض الأقصى (DC_LAT_MAX): 38.9955
- خط الطول الأدنى (DC_LNG_MIN): -77.1198
- خط الطول الأقصى (DC_LNG_MAX): -76.9094

بعد تحديد هذه الحدود، تم إنشاء قناع (mask) لتحديد الرحلات التي تقع نقطة انطلاقها أو وصولها خارج هذا الصندوق المحيط. وقد تبين أن **144,879** رحلة (ما يعادل 2.38% من إجمالي عدد الرحلات) تقع خارج هذه الحدود، نظراً لأن هذه الرحلات لا تمثل أنماط استخدام ضمن النطاق الجغرافي للمشروع، فقد تقرر التخلص منها من مجموعة البيانات لضمان تركيز التحليل على الرحلات ذات الصلة داخل واشنطن العاصمة.

3.6. معالجة مشكلة المعرفات المكررة للرحلات :

المشكلة: عند فحص عمود ride_id في مجموعة بيانات رحلات الدراجات اليوم، تبين وجود 123 قيمة مكررة لهذا المعرف، على الرغم من عدم وجود قيم مفقودة في هذا العمود، فإن تكرار ride_id يشير إلى وجود سجلات متطابقة أو شبه متطابقة لرحلات الدراجات أو سجلات خاطئة، مما قد يؤثر على دقة الإحصائيات والتحليلات الخاصة بالرحلات الفردية.

الحل المتبع: لمعالجة هذه المشكلة، تم إزالة السجلات المكررة بناءً على عمود ride_id، تم الاحتفاظ بالظهور الأول لكل ride_id فريد فقط، مما يضمن أن كل رحلة ممثلة بسجل واحد فقط في مجموعة البيانات، وبالتالي الحفاظ على سلامة ودقة البيانات للتحليلات اللاحقة.

3.7

4. هندسة سمات (Feature engineering) :

تعد هندسة السمات خطوة أساسية في عملية التنقيب عن البيانات، حيث يتم إنشاء متغيرات جديدة من البيانات الموجودة لتحسين قدرة النماذج التحليلية والتنبؤية على اكتشاف الأنماط والرؤى. في هذا المشروع، تركزت هندسة السمات على استخلاص معلومات زمنية ومكانية وحسابية إضافية من البيانات الخام.

4.1. استخلاص المكونات الزمنية من الرحلات :

لتحليل أنماط استخدام الدراجات بناءً على الوقت، تم فصل مكونات التاريخ والوقت من أعمدة started_at و ended_at في مجموعة بيانات الرحلات إلى أعمدة منفصلة.

4.2. حساب مدة الرحلة (Trip Duration) :

تم حساب مدة الرحلة بالدقائق (trip_duration_minutes) عن طريق طرح وقت بدء الرحلة (start_time) من وقت انتهائها (end_time) وتحويل الفرق إلى دقائق ، ومع ذلك عند فحص الإحصائيات الوصفية لعمود trip_duration_minutes، لوحظ وجود قيم سالبة، مما يشير إلى أخطاء في البيانات على سبيل المثال، كانت القيمة الدنيا لمدة الرحلة 12,692,040 - دقيقة، وهي قيمة غير منطقية.

الحل المتبع:

1. حساب trip_duration_minutes: تم حساب مدة الرحلة بالدقائق وتقريبها إلى منزلتين عشريتين لزيادة الدقة.
2. تحديد القيم غير الصالحة: تم تحديد 1158 سجلاً تحتوي على مدة رحلة سالبة أو صفرية، هذه السجلات لا تمثل رحلات صحيحة في سياق المشروع.
3. إزالة القيم غير الصالحة: تم إسقاط جميع الرحلات التي كانت مدتها أقل من أو تساوي الصفر من مجموعة البيانات.

بعد هذه المعالجة، أصبحت الإحصائيات الوصفية لعمود trip_duration_minutes منطقية، مما يؤكد أن جميع الرحلات المتبقية لها مدة إيجابية وصالحة للتحليل. هذه الخطوة ضرورية لضمان دقة أي تحليل يعتمد على مدة الرحلة، بما في ذلك حساب التكاليف والتنبؤ بالإيرادات.

4.3. حساب تكلفة الرحلة (Trip Cost) :

تم تصميم عملية حساب التكلفة لتأخذ في الاعتبار عدة عوامل معقدة، بما في ذلك نوع المستخدم (مشارك/عابر)، نوع الدراجة (كهربائية/كلاسيكية)، مدة الرحلة، وكون نقطة البدء أو النهاية تقع ضمن المنطقة التجارية المركزية (CBD) .

معالجة القيم الشاذة في التكلفة: بعد حساب التكاليف، كشفت الإحصائيات الوصفية لعمود trip_cost عن وجود قيم شاذة بشكل واضح، حيث بلغ الحد الأقصى للتكلفة حوالي 4.3 مليون دولار، وكان الانحراف المعياري مرتفعاً جداً (4,838) ، مما يشير إلى وجود عدد من الرحلات ذات التكاليف المبالغ فيها والتي لا تعكس الواقع وقد تؤثر سلباً على التحليلات.

الحل المتبع:

1. تحديد القيم الشاذة: تم تحديد الرحلات التي تزيد تكلفتها عن عتبة معينة (تم تحديدها بألف دولار 1000) كقيم شاذة مرتفعة، كما تم التحقق من عدم وجود قيم سالبة في التكلفة.
2. إزالة القيم الشاذة: تم إسقاط هذه الرحلات الشاذة من مجموعة البيانات، وقد تبين أن عدد هذه الرحلات الشاذة قليل جداً 40 رحلة فقط، ما يمثل نسبة 0.00% من إجمالي البيانات.

بعد إزالة هذه القيم الشاذة، أصبحت الإحصائيات الوصفية لعمود trip_cost أكثر واقعية ومنطقية، حيث أصبح متوسط التكلفة حوالي 3.78 دولار، وبلغت القيمة القصوى حوالي 79 دولار، مما يوفر مقياساً موثقاً للإيرادات المحتملة لكل رحلة.

4.4. تصنيف سعة المحطات :

لفهم أفضل لأنماط استخدام الدراجات وعلاقتها بحجم المحطة، كان من الضروري تصنيف المحطات بناءً على سعتها إلى فئات "صغيرة"، "متوسطة"، و"كبيرة"، التحدي يكمن في تحديد العتبات (Thresholds) المناسبة لهذه الفئات بطريقة منطقية وتكون قابلة للتعميم قدر الإمكان.

الاستكشاف الأولي لسعة المحطات:

- تم فحص الإحصائيات الوصفية لعمود CAPACITY في مجموعة بيانات المحطات (stations_df).
- أظهر توزيع السعة من خلال الرسوم البيانية التكرارية ومنحنى الكثافة المقدرة بالكيرنل KDE أن معظم المحطات تتركز حول سعة 15-19 دراجة، مع وجود عدد قليل من المحطات ذات السعة الكبيرة جداً.

النهج الأول: التصنيف بناءً على النسب المئوية (Quantiles)

الطريقة: تم استخدام النسب المئوية (Quantiles) لتقسيم البيانات إلى ثلاث فئات متساوية تقريباً في الحجم:

- الصغيرة (Small) (33%)
- المتوسطة (Average) (33% - 66%)
- الكبيرة (Large) (66%)
- تم تحديد العتبات بناءً على النسب المئوية 0.33 و 0.66 من توزيع السعة.
- **النتائج:**
 - صغيرة: 417 محطة
 - متوسطة: 255 محطة
 - كبيرة: 122 محطة

على الرغم من أن هذا النهج يضمن توزيعاً متوازناً للفئات بناءً على البيانات المتاحة، إلا أنه **خاص ببيانات واشنطن العاصمة فقط**. هذا يعني أن المحطات التي تُصنف "كبيرة" في واشنطن العاصمة قد لا تعتبر كبيرة في سياق عالمي أو في مدن أخرى، هذا النقص في التعميم جعل هذا النهج غير مثالي للتحليلات التي قد تتطلب مقارنات أو تطبيقاتاً في سياقات أوسع.

النهج الثاني: التصنيف بناءً على عتبات منطقية (Heuristic Thresholds)

الطريقة: تم اعتماد عتبات رقمية ثابتة بناءً على فهم عملي لسعة المحطات، مما يوفر تصنيفاً أكثر قابلية للتعميم وذات مغزى:

- الصغيرة: $Small \leq 15$
- المتوسطة: $15 < Average \leq 25$
- الكبيرة: $Large > 25$
- **النتائج:**
 - صغيرة: 417 محطة
 - متوسطة: 327 محطة
 - كبيرة: 50 محطة

تم تفضيل هذا النهج لأنه يوفر تصنيفاً أكثر منطقية وقابلية للتفسير، ولا يعتمد فقط على التوزيع الحالي للبيانات في واشنطن العاصمة، انه يعكس تمثيلاً أكثر دقة لأنواع المحطات بناءً على سعتها الفعلية، مما يجعله أكثر فائدة لاتخاذ القرارات التشغيلية وتحليلات الأداء، تم إضافة هذا التصنيف إلى مجموعة بيانات المحطات تحت عمود STATION_SIZE.

4.5. حساب المسافة إلى أقرب محطة نقل عام

لفهم مدى تأثير القرب من مراكز النقل العام على استخدام الدراجات، كان من الضروري حساب المسافة من نقاط بدء وانتهاء الرحلات إلى أقرب محطات المترو والحافلات الموكبة، هذه الميزة ستساعد في تحديد ما إذا كانت الدراجات تُستخدم كجزء من رحلات "الميل الأخير" (last-mile connectivity) "أو كوسيلة نقل أساسية من وإلى نقاط النقل الرئيسية.

خطوات الحساب:

1. **تحويل الإحداثيات الجغرافية (Projection):** نظراً لأن حساب المسافات الدقيقة يتطلب إحداثيات مستوية (planar coordinates) بدلاً من الإحداثيات الجغرافية (خط الطول/خط العرض)، تم تحويل جميع نقاط الانطلاق والانتهاج للرحلات، بالإضافة إلى إحداثيات محطات المترو والحافلات الموكبة، إلى نظام الإحداثيات المرجعية

EPSG:6933 هذا النظام هو نظام إسقاط (projected coordinate system) مناسب لحساب المسافات بالمتري بدقة.

2. حساب المسافة إلى أقرب محطة :

- لكل نقطة انطلاق رحلة، تم حساب المسافة الإقليدية (Euclidean distance) إلى أقرب محطة مترو وإلى أقرب محطة حافلات مكوكية.
 - وبالمثل، تم حساب نفس المسافات لكل نقطة انتهاء رحلة.
 - تم استخدام هيكل بيانات شجري متخصص (BallTree) لزيادة كفاءة البحث عن أقرب نقطة، خصوصاً مع الحجم الكبير لبيانات الرحلات.
3. **تحويل الوحدات**: تم تحويل جميع المسافات المحسوبة من المتر إلى **الكيلومتر** لسهولة التفسير.
4. **إنشاء الأعمدة الجديدة**: تم إضافة أربعة أعمدة جديدة إلى مجموعة بيانات الرحلات (trips_df) لتخزين هذه المسافات

على الرغم من معالجة القيم المفقودة وإزالة الرحلات خارج حدود واشنطن العاصمة في مراحل سابقة، لوحظ وجود عدد من الرحلات التي تحتوي على قيم مسافات "شاذة" إلى أقرب محطات المترو أو الحافلات، هذه المسافات كانت كبيرة، وتشير إلى أن بعض نقاط الرحلات، على الرغم من أنها ضمن حدود المدينة، لا ترتبط بشكل واقعي بشبكة الدراجات أو النقل العام ضمن مسافة المعروفة.

الحل المتبع:

1. **تحديد العتبات**: بناءً على تحليل التوزيعات الرسوم البيانية للمسافات (التي أظهرت ذيولاً طويلة لبعض القيم البعيدة)، تم تحديد عتبات عليا لكل نوع من المسافات كالتالي :
 - start_nearest_metro_distance
 - end_nearest_metro_distance
 - start_nearest_shuttle_distance
 - end_nearest_shuttle_distance
2. **إزالة القيم الشاذة**: تم إسقاط هذه الرحلات من مجموعة البيانات لضمان أن التحليلات اللاحقة تعتمد على مسافات منطقية وواقعية ضمن سياق التنقل الحضري.

4.6. تحديد ما إذا كانت الرحلة تمر بالمنطقة التجارية المركزية (CBD) :

تُعد المنطقة التجارية المركزية (CBD) نقطة جذب رئيسية للأنشطة الاقتصادية والاجتماعية في أي مدينة، لفهم أنماط استخدام الدراجات المتعلقة بالرحلات اليومية للعمل، أو التسوق، أو الترفيه في قلب المدينة، كان من الضروري إنشاء ميزة ثنائية (Binary Feature) توضح ما إذا كانت الرحلة تبدأ أو تنتهي ضمن هذه المنطقة الحيوية، هذه الميزة ستساعد في تحليل العلاقة بين استخدام الدراجات والمراكز الحضرية.

خطوات التنفيذ:

1. **تحميل وتجهيز بيانات المنطقة التجارية المركزية (CBD)** : تم تحميل بيانات المضلع الجغرافي للمنطقة التجارية المركزية.
2. **ضمان توحيد أنظمة الإحداثيات المرجعية (CRS)** : لضمان دقة التحليل المكاني (تحديد ما إذا كانت النقطة داخل المضلع)، تم التأكد من أن جميع الكائنات الجغرافية (نقاط بدء وانتهاء الرحلات، ومضلع CBD) تستخدم نفس نظام الإحداثيات المرجعية، تم تحويل نقاط الرحلات ومضلع CBD إلى نظام EPSG:6933، وهو نظام إسقاط مناسب للعمليات المكانية الدقيقة.
3. **التحقق من الاحتواء المكاني (Spatial Containment Check)** :
 - لكل رحلة، تم التحقق بشكل منفصل ما إذا كانت نقطة الانطلاق تقع داخل مضلع CBD
 - وبالمثل، تم التحقق ما إذا كانت نقطة الانتهاء تقع داخل مضلع CBD
4. **إنشاء الميزة الثنائية النهائية**: تم دمج النتائج لإنشاء عمود جديد باسم in_cbd هذا العمود يحمل قيمة True إذا كانت نقطة الانطلاق/أو نقطة الانتهاء للرحلة تقع ضمن المنطقة التجارية المركزية، وقيمة False بخلاف ذلك.

هذه الميزة ستكون بالغة الأهمية في فهم دور الدراجات في التنقل الحضري داخل وحول قلب المدينة.

4.7. حساب المسافة إلى المنطقة التجارية المركزية (CBD) وتصنيف القرب :

بالإضافة إلى معرفة ما إذا كانت الرحلة تبدأ أو تنتهي داخل المنطقة التجارية المركزية (CBD) ، من المهم فهم العلاقة المكانية للرحلات التي تنتهي بالقرب من هذه المنطقة، هذا يساعد في تحليل أنماط التنقل من وإلى ضواحي المنطقة التجارية، أو المناطق السكنية المجاورة، لتحقيق ذلك تم حساب المسافة من نقطة انتهاء كل رحلة إلى مركز المنطقة التجارية المركزية، ثم تصنيف الرحلات إلى "قريبة" أو "بعيدة" عن CBD .

خطوات التنفيذ:

1. **تحديد مركز المنطقة التجارية المركزية (CBD Centroid)** : تم تحديد النقطة المركزية (centroid) لمضلع المنطقة التجارية المركزية.
2. **تحويل الإحداثيات الجغرافية**: تم التأكد من أن إحداثيات نقطة انتهاء الرحلة ومضلع CBD والنقطة المركزية كلها مسقطة على نظام الإحداثيات EPSG:6933 ، لضمان دقة حساب المسافات بالمتر.
3. **حساب المسافة الإقليدية**: تم حساب المسافة الإقليدية (بالأمتار) من كل نقطة انتهاء رحلة إلى النقطة المركزية لل CBD، تم تخزين هذه المسافة في عمود جديد باسم distance_to_cbd_m.
4. **معالجة الرحلات التي تقع بالكامل داخل CBD**: نظراً لأن ميزة "المسافة إلى CBD " تهدف إلى قياس القرب لغير الرحلات الداخلية بالكامل، فقد تم تحديد الرحلات التي تبدأ وتنتهي كلاهما داخل المنطقة التجارية المركزية (المعلمة بواسطة start_in_cbd و end_in_cbd التي تم إنشاؤها سابقاً)، بالنسبة لهذه الرحلات، تم تعيين قيمة distance_to_cbd_m إلى null .

تحديد عتبة القرب (Thresholding) : بعد حساب المسافات، تم تحليل توزيع عمود distance_to_cbd_m باستخدام الإحصائيات الوصفية والرسوم البيانية التكرارية (Histograms) مع منحنيات الكثافة المقدرة بالكيرنل (KDE).

- أظهر التوزيع أن معظم الرحلات تنتهي على بعد بضعة كيلومترات من مركز CBD .
- لتحديد عتبة "القرب"، تم اختيار **الوسيط (Median)** لتوزيع المسافات كقيمة مناسبة، كانت قيمة الوسيط 2764 متر (أي حوالي 2.76 كيلومتر)، يعكس هذا الاختيار نقطة المنتصف في توزيع المسافات، حيث أن نصف الرحلات المتبقية تنتهي على مسافة أقل من هذه القيمة والنصف الآخر على مسافة أكبر.

إنشاء الميزة الثنائية "قريب من CBD " : بناءً على العتبة المختارة، تم إنشاء ميزة ثنائية جديدة باسم close_to_cbd. أما الرحلات التي تم تعيين distance_to_cbd_m لها ك null ، فإن close_to_cbd لهذه الرحلات تكون أيضاً Null.

ستمكن هذه الميزة من تحليل أنماط استخدام الدراجات للرحلات التي تخدم المنطقة التجارية المركزية ومحيطها القريب، مما يوفر رؤى حول استخدام الدراجات للتنقل الحضري اليومي.

4.8. تجميع المواقع الجغرافية باستخدام الهاش الجغرافي (Geohashing) :

تُعد نقاط الإحداثيات الجغرافية الخام (خطوط الطول والعرض) شديدة التناثر، مما يجعل من الصعب تحليل الأنماط المكانية العامة أو تجميع البيانات على مستوى الأحياء أو المناطق. لتسهيل التحليل المكاني على مستوى قطاعات جغرافية محددة، كان من الضروري استخدام تقنية الهاش الجغرافي (Geohashing) لتجميع المواقع في "قطاعات" قابلة للتحليل.

الهاش الجغرافي (Geohashing) : هي طريقة لترميز المواقع الجغرافية (خطوط الطول والعرض) في سلسلة نصية قصيرة، كلما زاد طول هذه السلسلة (عدد البتات)، زادت دقة الترميز وتصغر المساحة الجغرافية التي يمثلها الهاش، هذا يسمح بتمثيل المناطق الجغرافية المختلفة بمستويات مختلفة من التفصيل.

خطوات التنفيذ وتحديد عدد البتات المناسب:

1. **تقدير النطاق الجغرافي:** تم الأخذ في الاعتبار أن مساحة واشنطن العاصمة تبلغ حوالي 16 كم (شمال-جنوب) و 13 كم (شرق-غرب)، هذا التقدير الأولي يشير إلى أن دقة الهاش الجغرافي المناسبة للمدينة قد تتراوح بين 5 و 8 بتات.
2. **توليد الهاشات الجغرافية:** تم إنشاء أعمدة هاش جغرافي لنقاط بداية الرحلات (start_lat, start_lng) باستخدام مستويات دقة مختلفة (5، 6، 7، و 8 بتات).
3. **تقييم مستويات الدقة:** لتقييم الدقة المثلى، تم فحص مقياسين رئيسيين :
 - **عدد المناطق الفريدة (Unique Regions):**
 - دقة 5 بت 76: منطقة فريدة.
 - دقة 6 بت 989: منطقة فريدة.
 - دقة 7 بت 3595: منطقة فريدة.
 - دقة 8 بت 12398: منطقة فريدة.
 - **الوسيط لعدد الرحلات لكل هاش جغرافي (Median Trips per Geohash):**
 - دقة 5 بت 1761: رحلة لكل هاش.
 - دقة 6 بت 196: رحلة لكل هاش.
 - دقة 7 بت 7: رحلات لكل هاش
 - دقة 8 بت 2: رحلة لكل هاش.

القرار النهائي: بناءً على التقييم، تم اختيار **دقة 6 بتات** للهاش الجغرافي، توفر هذه الدقة توازناً ممتازاً بين تجزئة المناطق (989 منطقة فريدة) ووجود عدد كافٍ من الرحلات داخل كل قطاع (متوسط 196 رحلة لكل هاش)، مما يجعلها مناسبة للتحليل المكاني وتحديد الأنماط الإقليمية دون الإفراط في التجميع أو التناثر، تم تخزين هذه الميزة الجديدة في عمود باسم `geohash_sector`.

4.9. تقسيم القطاعات الجغرافية (Geohash Sectors) بناءً على حجم الرحلات اليومية :

بعد تجميع الرحلات في قطاعات جغرافية باستخدام الهاش الجغرافي، من المفيد تصنيف هذه القطاعات بناءً على حجم استخدامها (عدد الرحلات اليومية) إلى فئات مثل "عالية"، "متوسطة"، و"منخفضة"، هذا يساعد في تحديد المناطق ذات الطلب المرتفع على خدمات مشاركة الدراجات، وتلك التي تحتاج إلى تحسين أو زيادة في عدد المحطات.

الأساليب المقترحة: قدمت طريقتين محتملتين لتقسيم القطاعات:

1. **الكميات المئوية (Quantiles):** تقسيم القطاعات إلى ثلاث مجموعات متساوية بناءً على توزيع متوسط عدد الرحلات اليومية.
2. **تجميع K-Means:** استخدام خوارزمية التعلم الآلي لتجميع القطاعات في ثلاث مجموعات بناءً على تشابهها في متوسط عدد الرحلات اليومية.

التحليل والمقارنة:

- **الكميات المئوية :**
 - **المزايا:** بسيطة وسهلة التنفيذ، تضمن توزيعاً متساوياً للقطاعات بين الفئات.
 - **العيوب:** قد لا تعكس الاختلافات الحقيقية في حجم الاستخدام إذا كان التوزيع غير منتظم.
- **تجميع K-Means:**
 - **المزايا:** تأخذ في الاعتبار شكل توزيع البيانات، وقد تفصل المجموعات بشكل أفضل إذا كانت هناك فجوات واضحة بينها.
 - **العيوب:** أكثر تعقيداً، وقد تتأثر بالقيم المتطرفة، وقد لا تكون قابلة للتفسير بسهولة مثل الكميات المئوية.

الخطوات المتبعة :

1. **حساب متوسط عدد الرحلات اليومية لكل قطاع جغرافي :**

- تم تجميع البيانات حسب القطاع الجغرافي (geohash_p6) والتاريخ لحساب عدد الرحلات اليومية في كل قطاع.
- ثم تم حساب متوسط عدد الرحلات اليومية لكل قطاع.
- 2. **تطبيق طريقة التجميع المختارة :**
 - **الكميات المئوية:** تم حساب الكميات المئوية (الثلث والثلثين) لمتوسط عدد الرحلات اليومية، تم استخدام هذه القيم لتصنيف القطاعات إلى "رمادي" (منخفض)، "أصفر" (متوسط)، و "أحمر" (عالي).
 - **تجميع K-Means :** تم استخدام خوارزمية K-Means لتقسيم القطاعات إلى ثلاث مجموعات، تم تعيين تسميات "رمادي"، "أصفر"، و "أحمر" بناءً على متوسط عدد الرحلات في كل مجموعة (تم ترتيب المجموعات من الأدنى إلى الأعلى).
- 3. **دمج النتائج في مجموعة بيانات الرحلات :**
 - تم دمج تسميات مع مجموعة بيانات الرحلات الأصلية (trips_df)

الاختبار الأفضل هنا هو استخدام **تجميع K-Means** لأنه يأخذ في الاعتبار شكل توزيع البيانات وقد يفصل القطاعات بشكل أفضل.

سنتمكن الآن من تحليل أنماط استخدام الدراجات حسب حجم الاستخدام في كل قطاع جغرافي (أحمر، أصفر، رمادي). يمكن استخدام هذه المعلومات لتحديد المناطق ذات الطلب المرتفع، وتلك التي تحتاج إلى تحسين، وتخطيط توزيع المحطات بشكل أفضل.

4.10. تصنيف الظروف الجوية إلى فئات (Weather Segmentation) :

تحتوي مجموعة بيانات الرحلات على عمود "conditions" (الظروف الجوية) الذي يقدم وصفاً تفصيلياً ومتقلباً للطقس (مثل "غائم جزئياً"، "مُطر، غائم جزئياً"، "صافي"، إلخ)، هذا المستوى من التفصيل قد يكون صعباً للتحليل على نطاق واسع أو عند بناء نماذج تتطلب تصنيفات أوسع. لتسهيل فهم تأثير الظروف الجوية العامة على أنماط استخدام الدراجات، كان من الضروري تجميع هذه الظروف التفصيلية في فئات أوسع وأكثر دلالة.

خطوات التنفيذ:

1. **تحليل الظروف الحالية:** تم فحص القيم الفريدة وتوزيعها في عمود conditions لتحديد الأوصاف المختلفة للطقس.
2. **تعريف فئات الطقس:** تم تحديد ثلاث فئات رئيسية للطقس بناءً على الكلمات المفتاحية الموجودة في وصف الظروف :
 - **مُطر (rainy) :** تشمل أي ظروف تحتوي على كلمتي "مطر (rain)" أو "ثلج (snow)"
 - **غائم (cloudy) :** تشمل أي ظروف تحتوي على كلمتي "غائم (overcast)" أو "جزئياً غائم (partially cloudy)"
 - **مشمس (sunny) :** تشمل الظروف التي توصف بـ "صافي (clear)"
 - تم إضافة فئة "غير معروف (unknown)" كفئة افتراضية لأي وصف لا يندرج تحت الفئات المذكورة (على الرغم من أن جميع الظروف الحالية تم تغطيتها بالفئات الثلاث).
3. **تطبيق التصنيف:** تم إنشاء عمود جديد باسم weather_segment في مجموعة بيانات الرحلات (trips_df) ، وتم تطبيق دالة تقوم بتحويل الوصف النصي الأصلي للطقس إلى إحدى الفئات الثلاث المعروفة.

سيمكن هذا التصنيف المبسط من إجراء تحليلات أكثر فعالية حول كيفية تأثير الظروف الجوية العامة على عدد الرحلات، ومدة الرحلات، وسلوك المستخدمين بشكل عام.

4.11. إنشاء سلاسل زمنية للإيرادات اليومية حسب الظروف الجوية :

لفهم العلاقة بين الإيرادات اليومية وتأثير الظروف الجوية، من الضروري تجميع بيانات التكلفة (التي تمثل الإيرادات) على أساس يومي وربطها بفئة الظروف الجوية السائدة في ذلك اليوم. ستمكن هذه السلسلة الزمنية من تحليل الاتجاهات الموسمية واليومية للإيرادات وكيف تتأثر بتقلبات الطقس.

خطوات التنفيذ:

1. **توحيد تنسيق التاريخ:** تم التأكد من أن عمود ended_at في مجموعة بيانات الرحلات (trips_df) هو من نوع datetime، وذلك لضمان التعامل السليم مع مكونات التاريخ والوقت.
2. **استخلاص التاريخ اليومي:** تم استخلاص التاريخ فقط (بدون الوقت) من عمود ended_at وإنشاء عمود جديد باسم end_date، هذا يضمن تجميع الإيرادات على أساس يومي بغض النظر عن وقت انتهاء الرحلة الدقيق.
3. **تجميع الإيرادات اليومية حسب الظروف الجوية:**
 - تم تجميع مجموعة بيانات الرحلات (trips_df) حسب end_date التاريخ اليومي لانتهاء الرحلة، وweather_segment فئة الظروف الجوية.
 - تم حساب مجموع trip_cost (الإيرادات) لكل مجموعة، مما أدى إلى الحصول على الإيرادات اليومية لكل فئة طقس (مُطر، غائم، مشمس).
 - تم تخزين هذه البيانات المجمعة في إطار بيانات جديد باسم daily_income_weather.
4. **تجهيز البيانات للعرض:** تم تحويل عمود end_date في daily_income_weather مرة أخرى إلى نوع datetime لضمان التوافق مع أدوات تصور السلاسل الزمنية.

عرض السلاسل الزمنية (تنسيق طويل وتنسيق واسع): تم إنشاء تصورين لهذه السلاسل الزمنية لعرض العلاقة بين الإيرادات والظروف الجوية:

- **التنسيق الطويل: (Long Format)**
 - يعرض هذا التنسيق الإيرادات اليومية على محور Y والتاريخ على محور X، مع تمثيل كل فئة طقس (مُطر، غائم، مشمس) بخط منفصل بلون مختلف.
 - **الميزة:** يسهل مقارنة اتجاهات الإيرادات عبر فئات الطقس المختلفة في نفس الرسم البياني.
- **التنسيق الواسع: (Wide Format)**
 - في هذا التنسيق، يتم تحويل كل فئة طقس إلى عمود منفصل، بحيث يحتوي كل صف على التاريخ والإيرادات لكل فئة طقس في ذلك اليوم.
 - **الميزة:** يمكن أن يكون مفيداً لإجراء تحليلات إحصائية أو نمذجة حيث تكون كل فئة طقس متغيراً مستقلاً.

أيهما أفضل لمشكلتنا؟

بالنظر إلى هدف تحليل العلاقة بين الإيرادات والظروف الجوية، فإن **التنسيق الطويل (Long Format)** هو الأكثر ملاءمة والأكثر فعالية.

السبب:

- **المقارنة المباشرة:** يسمح التنسيق الطويل بمقارنة سهلة ومباشرة لاتجاهات الإيرادات وتأثيرها بالطقس على مدى فترة زمنية واحدة. يمكن رؤية كيف تتغير الإيرادات في الأيام الماطرة مقابل الأيام المشمسة أو الغائمة على نفس الرسم البياني، مما يسهل تحديد الأنماط الموسمية وتأثير الظروف الجوية المختلفة بشكل بصري.
- **سهولة التفسير:** يعد هذا التنسيق أكثر بديهية لتصور السلاسل الزمنية المتعددة التي تتقاسم محور زمن مشترك.

في المقابل، التنسيق الواسع قد يكون أقل وضوحاً بصرياً عند وجود العديد من الفئات، وقد تتداخل الخطوط أو يصعب تتبعها في رسم بياني واحد، لذلك للحصول على رؤى بصرية فورية حول تأثير الطقس على الإيرادات، يعتبر التنسيق الطويل هو الخيار الأمثل.

4.12 إنشاء سمات مساعدة للتحليل الاستكشافي للبيانات (EDA):

لتعزيز قدرتنا على استكشاف وفهم الأنماط الكامنة في بيانات رحلات الدراجات، تم إنشاء أربع سمات جديدة، صُممت هذه السمات لتبسيط الضوء على جوانب زمنية ومكانية محددة لسلوك المستخدمين.

1. ساعة الذروة rush_hour

- **الوصف:** ميزة ثنائية تشير إلى ما إذا كانت الرحلة قد بدأت خلال ساعات الذروة المعتادة في الصباح (من 7 صباحاً إلى 10 صباحاً) أو المساء (من 4 مساءً إلى 7 مساءً).
- **الهدف:** تهدف إلى تحديد ما إذا كان هناك اختلاف في أنماط استخدام الدراجات خلال فترات التنقل الأكثر ازدحاماً (الذهاب والإياب من العمل)، مما يساعد في فهم الاستخدام اليومي للركاب.

2. الفترة اليومية hour_segment

- **الوصف:** تصنيف زمني يقسم ساعات اليوم إلى أربع فترات أوسع: "الصباح" (من 5 صباحاً إلى 12 ظهراً)، "منتصف اليوم" (من 12 ظهراً إلى 5 مساءً)، "المساء" (من 5 مساءً إلى 9 مساءً)، و "الليل" (باقي الساعات).
- **الهدف:** تهدف إلى توفير نظرة عامة على كيفية تغير أنماط استخدام الدراجات على مدار اليوم (في فترات أكبر من الساعة الواحدة)، مما يكشف عن سلوكيات مختلفة مثل الرحلات الترفيهية مقابل رحلات العمل.

3. عطلة نهاية الاسبوع is_weekend

- **الوصف:** ميزة ثنائية تشير إلى ما إذا كانت الرحلة قد بدأت في يوم عطلة نهاية الاسبوع (السبت أو الأحد).
- **الهدف:** تهدف إلى كشف الاختلافات الجوهرية في أنماط استخدام الدراجات بين أيام العمل وعطلات نهاية الاسبوع، حيث يميل الناس لاستخدام الدراجات لأغراض ترفيهية أو اجتماعية في العطلات.

5. التحليل الاستكشافي للبيانات (Exploratory Data Analysis)

يمثل التحليل الاستكشافي للبيانات مرحلة حاسمة لفهم الأنماط الخفية، والعلاقات بين المتغيرات، وتحديد القيم المتطرفة ضمن مجموعات البيانات، في هذا المشروع سنستخدم مجموعة متنوعة من الرسوم البيانية والإحصائيات الوصفية لاستكشاف سلوك مستخدمي الدراجات وتأثير العوامل الزمنية والمكانية والبيئية على استخدامهم.

5.1 استخدام عينة للتحليل

نظراً للحجم الكبير لمجموعة البيانات (أكثر من 5 ملايين رحلة)، ولتحسين أداء وسرعة عمليات التحليل الاستكشافي للبيانات والتصورات، قمنا بأخذ عينة عشوائية من البيانات، مما يسمح بإجراء تحليلات سريعة وفعالة دون التأثير بشكل كبير على جودة الرؤى المستخلصة.

5.2 دراسة الحالات

A. دراسة حالة الدراجات و المحطات

B. دراسة مدة الرحلة

تُعد مدة الرحلة مقياساً أساسياً لفهم أنماط استخدام الدراجات، حيث يمكن أن تشير إلى الغرض من الرحلة (مثل التنقل اليومي أو المهام السريعة)، يهدف هذا القسم إلى استكشاف توزيع مدة الرحلات وتحديد الأنماط والقيم شاذة.

• رسم مخطط Histogram لتوزيع مدة الرحلة بالدقائق:

تم إنشاء مخطط Histogram لعرض توزيع مدة الرحلة بالدقائق، تم تحديد عدد الأنشطة (bins) في المخطط باستخدام قاعدة Freedman-Diaconis ، والتي تعد طريقة قوية لتحديد عرض الفئات بشكل ديناميكي بناءً على خصائص البيانات لتقديم أفضل تمثيل لتوزيعها.

التحليل والرؤى المستخلصة:

- **التركز الكبير على الرحلات القصيرة:** يُظهر المخطط بوضوح عموداً ضخماً يتركز بالقرب من الصفر، وتحديداً في النطاق من 0 إلى 20 دقيقة، هذا يشير إلى أن غالبية رحلات الدراجات قصيرة جداً في مدتها، وهو أمر نموذجي لأنظمة مشاركة الدراجات التي تُستخدم غالباً للتنقلات السريعة أو المهام اليومية في المسافات القصيرة.
- **وجود رحلات طويلة الأمد:** على الرغم من أن العدد قليل جداً، إلا أن المخطط يظهر وجود بعض الرحلات التي تمتد مدتها لتصل إلى 1440 دقيقة (أي ما يعادل يوم كامل).
- عند الفحص، تبين أن هناك 361 رحلة في مجموعة البيانات الكاملة تجاوزت مدتها 1440 دقيقة، هذا يشير إلى أن بعض الرحلات قد تستغرق أكثر من يوم واحد، وقد تمثل حالات استخدام خاصة (مثل الرحلات الطويلة جداً، أو الدراجات التي لم يتم إرجاعها في نفس اليوم لأسباب مختلفة).

- **رسم مخطط صندوقي (Box Plot) لتوزيع مدة الرحلة حسب نوع الدراجة:**
تم إنشاء مخطط صندوقي (Box Plot) لمقارنة توزيع مدة الرحلة بين أنواع الدراجات المختلفة (كلاسيكية وكهربائية)، يُظهر المخطط التوزيع الربيعي (Quartile Distribution) والقيم المتطرفة (Outliers) لكل نوع دراجة.

التحليل والرؤى المستخلصة:

- **تشابه كبير في مدة الرحلات النموذجية:** تُظهر الصناديق لكلا النوعين من الدراجات (الكلاسيكية والكهربائية) أنها شديدة الانضغاط وتقع بالقرب من علامة الـ 20 دقيقة على محور المدة، هذا يؤكد أن الغالبية العظمى من الرحلات، لكلا النوعين قصيرة جداً.
- **التوزيع المتناهي نحو اليمين (Right-Skewness):** يقع خط الوسيط (Median line) بالقرب جداً من الحد الأدنى للصندوق في كلا النوعين، هذا يشير إلى انحراف حاد نحو اليمين في توزيع البيانات، مما يعني أن معظم 50% من البيانات تتركز بالقرب من الطرف الأدنى، في حين أن الـ 25% الأعلى (حتى الربع الثالث) أكثر تشتتاً وتمتد نحو القيم الأطول.
- **وجود قيم متطرفة للرحلات الطويلة:** تمثل النقاط الخضراء فوق "الشعيرات (Whiskers)" القيم المتطرفة، والتي تشير إلى رحلات ذات مدة طويلة جداً، بعض هذه الرحلات تتجاوز بوضوح مدة 1440 دقيقة (يوم كامل).
- **مقارنة بين الدراجات الكلاسيكية والكهربائية:**
- **توزيع أكثر إحكاماً للدراجات الكهربائية:** يبدو أن الدراجات الكهربائية لديها توزيع مدة رحلات أكثر إحكاماً قليلاً مقارنة بالدراجات الكلاسيكية، هذا قد يشير إلى أن الدراجات الكهربائية تستخدم بشكل متكرر لرحلات قصيرة وفعالة، وربما لتغطية مسافات أطول في وقت أقل.
- **قيم متطرفة أكثر تطرفاً للدراجات الكلاسيكية:** تظهر الدراجات الكلاسيكية عدداً أكبر من القيم المتطرفة وبعضها يمتد إلى مدة أطول بكثير مقارنة بالدراجات الكهربائية، هذا يوحي بأن الدراجات الكلاسيكية تُستخدم أكثر للرحلات الطويلة، ربما بسبب عوامل مثل التكلفة (حيث قد تكون الدراجات الكهربائية أغلى للرحلات الطويلة)، أو قيود البطارية للدراجات الكهربائية، أو تفضيل المستخدمين للدراجات الكلاسيكية للرحلات الترفيهية الطويلة.

- **رسم مخطط صندوقي (Box Plot) لتوزيع مدة الرحلة حسب نوع العضوية:**
بطريقة مماثلة للمخطط السابق تم إنشاء مخطط صندوقي لمقارنة توزيع مدة الرحلة بين أنواع المشتركين.

التحليل والرؤى المستخلصة:

- **انحراف حاد نحو اليمين لكلا الفئتين:** على غرار الرسوم البيانية السابقة لمدة الرحلة، يُظهر مخطط الصندوق لكلا فئتي "المستخدمين العاديين (casual)" و "الأعضاء (member)" انحرافاً قوياً للغاية نحو اليمين، صندوق المخطط لكلا الفئتين مضغوط للغاية ويقع في الجزء السفلي من الرسم (قريباً من 0 دقيقة)، ويقع خط الوسيط عملياً فوق الربع الأول (Q1)، هذا يؤكد أن الغالبية العظمى من الرحلات لكلا النوعين من المستخدمين قصيرة جداً.
- **مدة رحلة أطول قليلاً للمستخدمين العاديين:** على الرغم من أن مدة الرحلات النموذجية قصيرة لكلا الفئتين، إلا أن الوسيط (والربع الأول/الثالث) للمستخدمين العاديين يبدو أعلى قليلاً (أو على الأقل، الصندوق أقل انضغاطاً بشكل هامشي) من تلك الخاصة بالأعضاء، هذا يشير إلى أن الرحلة "النموذجية" للمستخدم العادي أطول قليلاً من رحلة العضو، حتى لو كانت كلتا قصيرتين نسبياً.
- **وجود قيم متطرفة للرحلات الطويلة لكلا الفئتين:** تظهر لكلا الفئتين بوضوح عدداً كبيراً من الرحلات "المتطرفة" (النقاط الخضراء) التي تمتد إلى ما هو أبعد بكثير من الصندوق والشعيرات الرئيسية، مما يشير إلى أن الرحلات الطويلة جداً تحدث لكلا النوعين من المستخدمين.
- **كثافة أعلى للرحلات المتطرفة الطويلة للمستخدمين العاديين:** يمتلك المستخدمون العاديون كثافة أعلى بكثير من الرحلات المتطرفة الطويلة جداً، بينما يمتلك الأعضاء أيضاً رحلات متطرفة طويلة، إلا أنها أقل عدداً وبشكل عام لا تصل إلى نفس المدد القصوى التي يصل إليها المستخدمون العاديون، هذا يشير إلى أن المستخدمين العاديين قد يستخدمون الدراجات لرحلات أطول بشكل متقطع، ربما لأغراض ترفيهية أو سياحية، في حين أن الأعضاء يفضلون الرحلات القصيرة والمنتظمة التي تتناسب مع الاشتراكات.

كشف هذا الجزء من التحليل الاستكشافي للبيانات عن رؤى أساسية حول سلوك استخدام الدراجات في واشنطن العاصمة، بشكل عام، أظهرت الدراسة أن الغالبية العظمى من الرحلات، بغض النظر عن نوع الدراجة أو العضوية، تكون قصيرة جداً. ومع ذلك، لاحظنا اختلافات دقيقة تشير إلى أن المستخدمين العاديين يميلون إلى إجراء رحلات أطول قليلاً وقد يظهرون أنماط استخدام أكثر تنوعاً مقارنة بالأعضاء، كما حدد التحليل وجود عدد قليل من الرحلات طويلة الأمد للغاية، والتي تتركز بشكل خاص في وسط العاصمة وتنتشر عبر مجموعة واسعة من المحطات، مما يسلط الضوء على تحديات تشغيلية وفرص محتملة لأنماط استخدام متخصصة.

C. الدراسة المالية

D. التحليل المكاني لرحلات الدراجات

يهدف هذا القسم من التحليل الاستكشافي للبيانات إلى فهم الأنماط المكانية لاستخدام الدراجات، وتحديد المناطق ذات الكثافة العالية للرحلات، وكيف ترتبط هذه الأنماط بالتقسيمات الجغرافية المختلفة للمدينة.

- عرض مخطط حراري جغرافي لعدد الرحلات :
تم إنشاء مخطط حراري جغرافي لعرض كثافة نقاط بداية ونهاية الرحلات التي تقع داخل أو بالقرب من المناطق السكنية المحددة في واشنطن العاصمة، هذا المخطط يساعد في تحديد المناطق السكنية النشطة بشكل خاص فيما يتعلق باستخدام الدراجات، وتوضيح كيفية تفاعل الرحلات مع هذه المناطق.

خطوات التنفيذ:

1. تحميل وتوحيد بيانات المناطق السكنية والتأكد من أنها في نظام الإحداثيات المرجعية EPSG:4326 .
2. تم تحديد جميع نقاط بداية ونهاية الرحلات (من العينة المختارة) التي تقع داخل حدود المناطق السكنية.
3. تجميع النقاط للعرض.
4. تم إضافة حدود المناطق السكنية وحدود المنطقة التجارية المركزية (CBD) كطبقات إلى الخريطة لتقديم سياق مكاني أوضح.
5. إنشاء الخريطة الحرارية.

التحليل والرؤى المستخلصة:

- **تطابق عال مع المناطق السكنية:** الملاحظة الأبرز هي أن الكثافة العالية جداً لرحلات الدراجات تتوافق بشكل كبير وتتبع حدود المناطق السكنية، هذا يشير إلى أن نظام مشاركة الدراجات يُستخدم بكثافة لأغراض ترتبط مباشرة بالمناطق السكنية، سواء كانت رحلات تبدأ منها، أو تنتهي فيها، أو تمر عبرها.
- **تركز الاستخدام في مناطق محددة:** يظهر أعلى تركيز للرحلات ضمن المناطق السكنية في الأجزاء الوسطى والشمالية من واشنطن العاصمة، هذا يوحي بأن هذه المناطق هي الأكثر نشاطاً في استخدام الدراجات الهوائية السكنية.
- **مناطق ذات كثافة أقل:** بينما النمط العام كثيف، توجد بعض المناطق داخل الحدود السكنية تظهر كثافة أقل من الرحلات، قد تشير هذه المناطق إلى وجود عدد أقل من المحطات، أو خصائص ديموغرافية مختلفة، أو حاجة أقل لخدمات مشاركة الدراجات، وهي نقاط تستدعي المزيد من الاستكشاف في المستقبل.

- عرض مخطط شريطي لتوزيع فئات القطاعات الجغرافية (Geohash Sectors)
تم إنشاء مخطط شريطي (Bar Chart) لتوضيح توزيع عدد الرحلات عبر القطاعات الجغرافية المختلفة (الممثلة بواسطة geohash_p6) يهدف هذا المخطط إلى تحديد القطاعات الجغرافية الأكثر نشاطاً والتعرف على مدى تباين حجم النشاط بينها.

خطوات التنفيذ:

1. تم حساب عدد الرحلات لكل قطاع جغرافي باستخدام البيانات المأخوذة من العينة.
2. تم فرز القطاعات تنازلياً حسب عدد الرحلات لتسهيل تحديد القطاعات الأكثر نشاطاً بصرياً.
3. رسم المخطط الشريطي

التحليل والرؤى المستخلصة:

- **تركز النشاط في عدد قليل من القطاعات:** تظهر الأعمدة القليلة الأولى على اليسار أنها أعلى بكثير من البقية، حيث يقترب أطول عمود من 800 رحلة، بينما العديد من الأعمدة الأخرى تقل عن 50 رحلة، يشير هذا بوضوح إلى أن نشاط مشاركة الدراجات يتركز بشدة في عدد قليل من المناطق ذات الطلب المرتفع.
- **تحديد المناطق الأكثر طلباً:** يتيح الرسم البياني تحديد القطاعات الجغرافية ذات الطلب الأعلى بسرعة، من المرجح أن تتوافق هذه المناطق مع مناطق ذات كثافة سكانية عالية، أو محاور نقل رئيسية، أو مناطق جذب سياحي شهيرة، أو مناطق تجارية/مكتبية كثيفة.
- **تأكيد النقاط الساخنة:** عند مقارنة هذه القطاعات الأعلى نشاطاً بالخريطة الحرارية للمناطق السكنية من المهمة 1، نجد تأكيداً على أن هذه القطاعات تقع ضمن "النقاط الساخنة" التي تظهر كثافة عالية من الرحلات في المناطق السكنية، هذا يعزز الفهم بأن الاستخدام يتركز حول المراكز الحضرية الرئيسية والمناطق السكنية المتصلة بها جيداً.

• عرض مخططات Histogram لتوزيع المسافة عن المنطقة التجارية الرئيسية (CBD) ، وأقرب محطة مترو، وأقرب محطة حافلات:

للمزيد من التعمق في الأنماط المكانية، تم إنشاء ثلاثة مخططات Histogram منفصلة، تهدف هذه المخططات إلى تحليل توزيع الرحلات بناءً على مسافتها من نقاط اهتمام رئيسية: المنطقة التجارية المركزية (CBD) ، وأقرب محطة مترو، وأقرب محطة حافلات، تم استخدام قاعدة Freedman–Diaconis لتحديد عدد مناسب من الأنشطة (bins) لكل مخطط، مما يضمن تمثيلاً دقيقاً للتوزيع.

خطوات التنفيذ:

1. **حساب عدد الـ Bins:** تم تعريف دالة `freedman_diaconis_bins` لحساب العدد الأمثل للفئات (nbins) لكل توزيع بناءً على بيانات المسافات.
2. **تحويل المسافة إلى كيلومترات:** تم تحويل عمود `distance_to_cbd_m` من الأمتار إلى الكيلومترات (`distance_to_cbd_km`) لتسهيل القراءة والتفسير.
3. **رسم مخططات Histogram:** تم إنشاء ثلاثة مخططات Histogram لكل من :
 - المسافة إلى المنطقة التجارية المركزية.
 - المسافة إلى أقرب محطة مترو.
 - المسافة إلى أقرب محطة حافلات.

التحليل والرؤى المستخلصة:

1. توزيع المسافة إلى المنطقة التجارية المركزية (CBD)

- **التركيز القوي حول الـ CBD:** تُظهر الغالبية العظمى من الرحلات (ممثلة بالأعمدة الأعلى) أنها تبدأ أو تنتهي على مسافة تتراوح بين 0 و 4 كيلومترات من المنطقة التجارية المركزية، يبلغ ذروة التكرار حوالي 2.2 كيلومتر.
- **الدلالة:** هذا يؤكد الدور المحوري للمنطقة التجارية المركزية كمركز رئيسي لنشاط مشاركة الدراجات، يشير هذا بقوة إلى أن النظام يخدم في المقام الأول التنقلات القصيرة ("الميل الأخير" أو "الميل الأول") أو التنقلات داخل الـ CBD نفسها.

2. توزيع المسافة إلى أقرب محطة مترو

- **توزيع منحرف للغاية:** يعرض المخطط توزيعاً منحرفاً بشدة، حيث يظهر تكراراً مرتفعاً للغاية للرحلات التي تنطلق أو تنتهي بالقرب جداً من محطات المترو، ينخفض عدد الرحلات بشكل حاد مع زيادة المسافة عن محطة المترو.
- **الدلالة:** تُبرز هذه الرؤية أهمية وضع محطات مشاركة الدراجات على مسافة قريبة جداً من مداخل محطات المترو، هذا يعظم تكامل النظام مع شبكة النقل العام ويلبي احتياجات الركاب بفعالية، حيث تُستخدم الدراجات كوصلة أساسية للتنقل من وإلى المترو.

3. توزيع المسافة إلى أقرب محطة حافلات

- **نمط مشابه لمحطات المترو:** على غرار مخطط محطات المترو، يُظهر هذا المخطط توزيعاً منحرفاً للغاية، مع أعلى تكرار للرحلات التي تحدث بالقرب جداً من محطات الحافلات.

- **الدلالة:** يشير هذا إلى سلوك مستخدم ثابت يتمثل في استخدام الدراجات لسد الفجوات القصيرة للوصول إلى نقاط النقل العام الثابتة (سواء كانت مترو أو حافلات)، هذا يؤكد دور الدراجات كحل للتنقل متعدد الوسائط (multimodal transportation)، حيث تكمل شبكة الحافلات والمترو.

- عرض مخطط شريطي لتوزيع الرحلات بناءً على ما إذا كانت داخل المنطقة التجارية المركزية (CBD) بالكامل أم خارجها

لتحليل مدى أهمية المنطقة التجارية المركزية (CBD) كوجهة أو نقطة انطلاق للرحلات، تم تصنيف الرحلات إلى فئتين: "داخل CBD بالكامل" إذا بدأت وانتهت داخل الـ CBD أو "خارج CBD" إذا كانت نقطة البداية أو النهاية، أو كلاهما، خارج الـ CBD، تم عرض هذا التوزيع باستخدام مخطط شريطي.

خطوات التنفيذ:

1. **تصنيف الرحلات:** تم تعريف دالة classify_trip تقوم بتصنيف كل رحلة في العينة إلى داخل الـ CBD إذا كانت كل من نقطة البداية والنهاية داخل الـ CBD، وإلا يتم تصنيفها خارج الـ CBD.
2. **تطبيق التصنيف:** تم تطبيق هذه الدالة على إطار البيانات sampled_df لإنشاء عمود جديد cbd_trip_type.
3. **حساب التكرارات:** تم حساب عدد الرحلات لكل فئة (cbd_trip_type) في العينة وعرضها في مخطط شريطي.
4. **حساب النسبة المئوية للمجموعة الكاملة:** تم تطبيق نفس منطق التصنيف على مجموعة البيانات الكاملة (trips_df) لحساب التوزيع النسبي للرحلات داخل وخارج الـ CBD للحصول على رؤية شاملة.

التحليل والرؤى المستخلصة:

- **الغالبية العظمى خارج الـ CBD:** تُظهر النتائج أن الغالبية العظمى من الرحلات (أكثر من 93% في مجموعة البيانات الكاملة) تندرج ضمن فئة "خارج CBD"، هذا يشير إلى أن نظام مشاركة الدراجات يخدم بشكل أساسي منطقة جغرافية أوسع تتجاوز المنطقة التجارية المركزية الأساسية، أو يسهل الاتصالات من وإلى الـ CBD بدلاً من التركيز حصراً على التنقلات داخلها.
- **حركة أقل داخل الـ CBD بالكامل:** عدد أقل بكثير من الرحلات (حوالي 6.26% في مجموعة البيانات الكاملة) صُنِّفت على أنها "داخل CBD بالكامل"، هذا يشير إلى أنه على الرغم من وجود بعض الحركة الداخلية في الـ CBD باستخدام الدراجات، إلا أنها تمثل نسبة صغيرة جداً من إجمالي الرحلات.
- **دعم حل "الميل الأول/الأخير":** يشير هذا التوزيع بقوة إلى أن الوظيفة الأساسية لخدمة مشاركة الدراجات لا تقتصر على التنقلات السريعة داخل الـ CBD نفسها، بل تتمحور حول تسهيل التنقلات اليومية، المهام، أو الرحلات الترفيهية التي تربط الـ CBD بأجزاء أخرى من المدينة، أو تعبرها، يتوافق هذا مع الرؤى المستخلصة من مخططات توزيع المسافة، حيث غالباً ما تربط الرحلات المستخدمين بمحاور النقل الرئيسية (مثل محطات المترو والحافلات) أو المناطق السكنية التي قد تقع خارج الحدود الصارمة للمنطقة التجارية المركزية، إن حقيقة أن أكثر من 93% من جميع الرحلات تتضمن مناطق خارج الـ CBD تعزز بقوة فكرة أن مشاركة الدراجات تُستخدم بشكل أساسي كحل "للميل الأول/الأخير".

- **عرض مخطط شريطي لتوزيع الرحلات التي مرت عبر المنطقة التجارية المركزية (CBD) حسب نوع الدراجة ونوع العضوية**

لتحليل أنماط استخدام الدراجات داخل وحول المنطقة التجارية المركزية (CBD) بشكل أعمق، تم إنشاء مخطط شريطي يوضح عدد الرحلات التي مرت بالـ CBD (أي بدأت أو انتهت داخلها) مصنفة حسب نوع الدراجة (كلاسيكية أو كهربائية) ونوع العضوية (عضو أو عادي).

خطوات التنفيذ:

1. **تصفية الرحلات التي مرت بالـ CBD:** تم تصفية العينة (sampled_df) لتشمل فقط الرحلات التي كانت نقطة بدايتها أو نقطة نهايتها داخل المنطقة التجارية المركزية (start_in_cbd == 1 أو end_in_cbd == 1).
2. **تجميع البيانات:** تم تجميع الرحلات المصفاة حسب نوع الدراجة و نوع العضوية وحساب عدد الرحلات لكل مجموعة فرعية.
3. **رسم المخطط الشريطي**

4. **حساب النسبة المئوية لإجمالي الرحلات:** تم تطبيق نفس منطق التصفية والتجميع على مجموعة البيانات الكاملة (trips_df) لتحديد النسبة المئوية للرحلات التي تمر عبر ال CBD من إجمالي الرحلات، تبين أن **29%** من جميع الرحلات تمر عبر المنطقة التجارية المركزية.

التحليل والرؤى المستخلصة:

- **الدور المحوري لل CBD:** النسبة المئوية العالية (29%) لجميع رحلات مشاركة الدراجات التي تمر عبر المنطقة التجارية المركزية تشير إلى دورها الحاسم كنقطة انطلاق أو وجهة، أو نقطة عبور لجزء كبير من مستخدمي الدراجات.
- **هيمنة الأعضاء في رحلات ال CBD:** بالنسبة لكل من الدراجات الكلاسيكية والكهربائية، يمثل الأعضاء (الاعمدة البرتقالية/الحمراء) باستمرار عدداً أعلى بكثير من الرحلات التي تمر عبر ال CBD مقارنة بالمستخدمين العابرين (الاعمدة الزرقاء)، هذا يوحي بأن الركاب المنتظمين أو المستخدمين المتكررين (الأعضاء) هم أكثر عرضة لاستخدام مشاركة الدراجات للتنقلات التي تشمل المنطقة التجارية المركزية.
- **تفضيل الدراجات الكلاسيكية في ال CBD:** يُظهر الرسم البياني أن كلا من المستخدمين العابرين والأعضاء يستخدمون الدراجات الكلاسيكية بشكل متكرر لرحلات ال CBD أكثر من الدراجات الكهربائية، هذا قد يعكس تفضيلاً عاماً للدراجات الكلاسيكية، أو ربما يكون مؤشراً على توفرها بشكل أكبر في المناطق المكتظة، أو أنها مفضلة لأنواع معينة من الرحلات القصيرة داخل المناطق الحضرية الكثيفة.
- **التوصيات التشغيلية:** نظراً لأن الأعضاء هم المستخدمون الأساسيون للرحلات المتعلقة بال CBD، فإن الاستراتيجيات التي تهدف إلى الاحتفاظ بالعضوية وتنميتها، وضمان توفر الدراجات الكافي داخل وحول ال CBD، تعتبر حاسمة. كما أن الطلب على كلا النوعين من الدراجات في رحلات ال CBD يعني أن إدارة الأسطول يجب أن تأخذ في الاعتبار توزيعاً متوازناً لتلبية تفضيلات كل من الأعضاء والمستخدمين العابرين.

• دراسة الارتباط بين فئة البعد عن المنطقة التجارية الرئيسية (CBD) ونوع العضوية باستخدام تحليل Chi-Square

لتقييم ما إذا كان هناك ارتباط إحصائي ذو دلالة بين قرب الرحلة من المنطقة التجارية المركزية (CBD) ونوع عضوية المستخدم (عضو أو عادي)، تم إجراء اختبار Chi-Square هذا التحليل يساعد في فهم ما إذا كان نمط استخدام ال CBD يختلف بشكل كبير بين الأعضاء والمستخدمين العاديين.

خطوات التنفيذ:

1. إنشاء جدول التوافق (Contingency Table): تم إنشاء جدول تكراري يوضح عدد الرحلات لكل مزيج من المتغيرات قريب من ال CBD أو لا و عضو أو عادي .
2. إجراء اختبار Chi-Square: تم تطبيق وظيفة chi2_contingency من مكتبة scipy.stats على جدول التوافق لحساب قيمة إحصائية، قيمة p، درجات الحرية، والتكرارات المتوقعة.
3. تفسير النتائج: تم تفسير قيمة p لتحديد ما إذا كان الارتباط الملاحظ ذو دلالة إحصائية عند مستوى دلالة 0.05.

التحليل والرؤى المستخلصة:

- **دلالة إحصائية قوية:**
 - أظهرت النتائج قيمة χ^2 إحصائية عالية جداً (1504.37) وقيمة P-value تساوي 0.0، وهي أقل بكثير من مستوى الدلالة التقليدي 0.05.
 - الاستنتاج: هذا يشير بوضوح إلى وجود ارتباط ذو دلالة إحصائية قوية بين ما إذا كانت الرحلة قريبة من المنطقة التجارية المركزية ونوع عضوية المستخدم، بمعنى آخر، نمط استخدام ال CBD لا يتوزع عشوائياً بين الأعضاء والمستخدمين العاديين.
- **الفروقات النسبية في الاستخدام:**
 - على الرغم من الدلالة الإحصائية القوية (والتي غالباً ما تكون متوقعة مع أحجام العينات الكبيرة جداً)، فإن الفروق النسبية في الجدول التكراري على سبيل المثال، نسبة رحلات "قريب من ال CBD" بين الأعضاء والعابرين (ليست دراماتيكية).
 - يشير الجدول إلى أن كلا من الأعضاء والمستخدمين العابرين لديهم ما يقرب من نصف رحلاتهم (أو نقاط بدايتها/نهايتها) قريبة من المنطقة التجارية المركزية.
 - الدلالة: هذا يعني أن بينما يوجد ارتباط إحصائي، فإن الفرق العملي في النسبة المئوية للرحلات القريبة من ال CBD بين الفئتين قد لا يكون كبيراً بالقدر الذي توحى به قيمة χ^2 وحدها، كلا النوعين من المستخدمين يستفيدون من الدراجات للتنقل في المنطقة المحيطة بال CBD، ولكن قد تكون الأسباب أو الأنماط الدقيقة مختلفة.

كشفت الدراسة المكانية أن استخدام الدراجات يتركز بشكل كبير في المناطق السكنية والمراكز الحضرية الرئيسية، مع دور محوري للمنطقة التجارية المركزية كنقطة ربط للرحلات، حتى لو كانت معظم الرحلات تمتد خارج حدودها الصارمة، تؤكد هذه الأنماط الجغرافية أن النظام يُستخدم بشكل أساسي كحل "للميل الأول/الأخير" لتسهيل التنقل من وإلى محاور النقل، وأن الأعضاء هم اللاعبون الرئيسيون في هذا النوع من التنقل الحضري.

E. دراسة تأثير حالة الطقس

5.3.

6.