أكاديمية طويق
TUWAIQ ACADEMY

معسكر علم البيانات و تعلم الآلة

13 -11 - 2022

أكاديمية طويق
TUWAIQ ACADEMY

نبذة عن المدرب

أكاديمية طويق
TUWAIQ ACADEMY

محتوى المعسكر

اليوم	الأسبوع الأول Getting Started	الأسبوع الثاني Data Analysis and Visualization	الأسبوع الثالث Machine Learning	الأسبوع الرابع EDA & FE in Action	الأسبوع الخامس Modeling Interpretation in Action	الأسبوع السادس Final Project
الأحد	Intro to DS	NumPy	Intro to ML	DS Knowledge Catalog	Models Families: Distance & Time Series	Final Project
الإثنين	Git & Github	Pandas	Supervised ML	EDA1: Univariate & Multivariate Analysis	Models Evaluation: Regression & Classification	Final Project
الثلاثاء	Python Review	Matplotlib	Supervised ML	**EDA2: Association Analysis & Hypothesis Construction**	Optimization Techniques	Final Project
الأربعاء	Python Review	Seaborn	Unsupervised ML	Features Engineering: Scaling, Merging & Discretization	NLP and Text Mining Basics	Final Project
الخميس	Python Review	Plotly	Unsupervised ML	Models Families: Continuous & Categorical	Neural Networks Basics	Presentation

**ملاحظة: قد تتغير المواضيع أو أوقات طرحها بناء على تقدم الطلاب.

مرحلة استكشاف البيانات EDA؟

ما هي مرحلة تحليل البيانات الاستكشافي؟

هي مرحلة يتم فيها التعرف على البيانات، وصفها، وحالتها، ومن ثم تنظيفها، وتُوجه الطريق للوصول إلى بيانات يُعتمد عليها في اتخاذ وتوجيه القرارات

الهدف من مرحلة تحليل البيانات الاستكشافي

لأننا نحاول في هذه المرحلة فهم البيانات والوصول إلى نسخة منقحة يعتمد عليها، إلا أن الطريق للوصول لهذه النسخة يُحدده الهدف من العمل على هذ المنتج.

تخدم هذه المرحلة الأهداف التالية:

1. الحصول على مرئيات عن البيانات ومشاركتها
2. فهم الهيكل الأساسي للبيانات
3. استخراج الخواص المهمة

الهدف من مرحلة تحليل البيانات الاستكشافي

لأننا نحاول في هذه المرحلة فهم البيانات والوصول إلى نسخة منقحة يعتمد عليها، إلا أن الطريق للوصول لهذه النسخة يُحدده الهدف من العمل على هذ المنتج.

تخدم هذه المرحلة الأهداف التالية:

4. الكشف عن القيم والحالات الشاذة
5. اختبار الافتراضات التي تم وضعها في مرحلة فهم المشكلة
6. توجيه الوصول إلى النتيجة النهائية

خطوات تحليل البيانات

1. تحديد المتطلبات من البيانات

تجيب هذه الخطوة على الأسئلة التالية:

- ما هو حجم البيانات؟
- هل تخدم أنواع البيانات الحالية النتيجة التي نهدف الوصول لها؟
- تحديد مصادر البيانات وكيف نشأت هذه الحقائق

2. تجميع وتنظيم البيانات للتحليل

تجيب هذه الخطوة على الأسئلة التالية:

- ما هي البيانات الأخرى التي أستطيع إضافتها لتعزز الهدف من العمل؟
- من أين سنقوم بجمعها وماهي التقنيات التي نحتاجها؟
- من سيقوم بجمعها؟

3. فحص صحة البيانات

تجيب هذه الخطوة على الأسئلة التالية:

- هل هناك حقائق فارغة؟ كيف سأتعامل معها؟
- هل هناك حقائق شاذة أو مغلوطة؟ كيف سأتعامل معها؟
- هل هناك بيانات مكررة؟

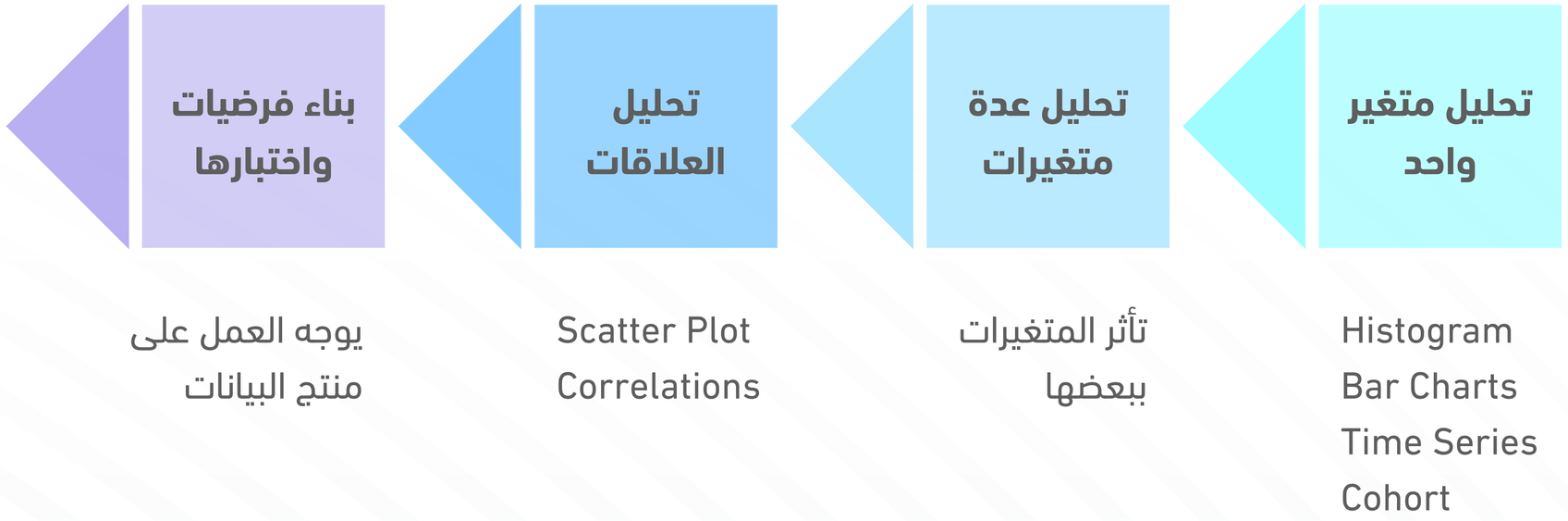
4. تحليل العلاقات في البيانات

تجيب هذه الخطوة على الأسئلة التالية:

- ما مدى تأثير كل عامود أو خاصية على النتيجة النهائية؟
- ماهي الأعمدة التي ستستمر للمرحلة التالية؟

أنواع تحليل البيانات

هناك عدة أنواع للتحليل الوصفي للبيانات



أكاديمية طويق
TUWAIQ ACADEMY

تحليل العلاقات correlation

العلاقات Correlations

هي أحد أمثلة التحليلات ذات متغيرين و**تحسب العلاقة بين متغيرين عدديّين**

العلاقات للمتغيرات العددية

معامل الارتباط **بيرسون**

$$r = \frac{\sum_{i=1}^{n}(x_i - x)(y_i - y)}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \overline{y})^2\right]}}$$

- قيمة العلاقة تنحصر بين ١ و -١

- ممكن نستخدم حساب العلاقة لتتعرف على تأثير المدخلات على المخرج، ومقارنة نتائج حجم التأثير

العلاقات للمتغيرات المصنفة

تجمهر البيانات في تقاطع معين بين التصنيفات

التعرف من العلاقات بين متغير مصنف عن متغير عددي

الهدف من هذه الخطوة هي معرفة ما إذا كانت هناك روابط بين المتغيرين

إذا كان المتغير الوصفي متأثرًا تأثر ملحوظا بالمتغير العددي بما يعني أن التصنيفات تنتمي إلى مجال متقارب من الأعداد

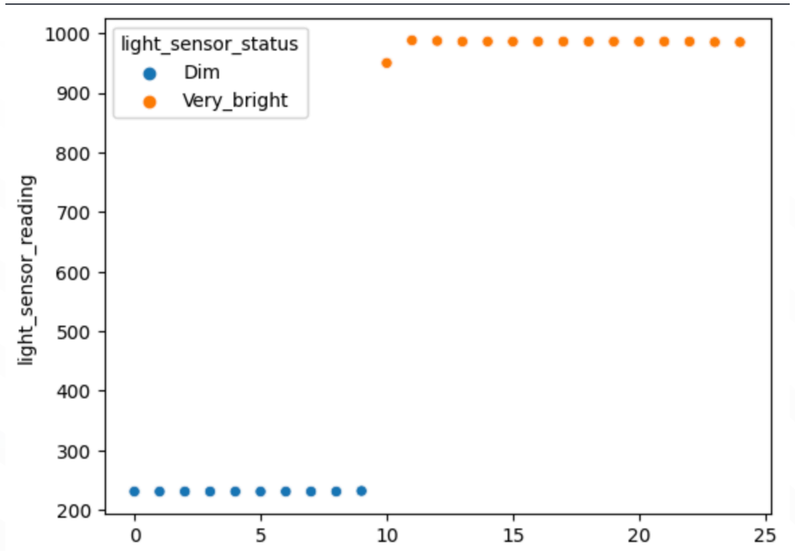
التعرف من العلاقات بين متغير مصنف عن متغير عددي

مثال: حساس ضوئي يحسب درجة الضوء (متغير عددي) وله حالتين (تصنيف) إما مظلم أو مضيء، ما هي الخطوات لحساب العلاقة بينهم؟

درجة الضوء	حالة الحساس
200	مظلم
897	مضيء
201	مظلم
899	مضيء

التعرف من العلاقات بين متغير مصنف عن متغير عددي

مثال: حساس ضوئي يحسب درجة الضوء (متغير عددي) وله حالتين (تصنيف) إما مظلم أو مضيء، ما هي الخطوات لحساب العلاقة بينهم؟



مصدر بيانات المثال: shorturl.at/mqHNZ

للتسليم

	Categorical	Continuous
Categorical	Lambda, Corrected Cramer's V	Point Biserial, Logistic Regression
Continuous	Point Biserial, Logistic Regression	Spearman, Kendall, Pearson

اختيار طريقة لإثبات علاقة بين متغيرين وشرحها

أكاديمية طويق
TUWAIQ ACADEMY

تحليل الروابط بين قيم المتغيرات
Association Rules

تحليل الروابط بين قيم التصنيف الواحد

هذا النوع من التحليل يُعد تحليل سلوكي، و**تجيب هذه الخطوة على أهم تساؤل:** هل التصنيفات تتقاطع بكثرة؟

- يساعد فهم الروابط في منتجات لبيانات التي تركز على التوصيات، مثل:

  * اقتراح شراء منتج س عند إضافة منتج ع إلى السلة
  * سبب ظهور اقتراح منتج س هو كثرة الطلب على المنتجين بنفس عملية الشراء

أمثلة

أكاديمية طويق
TUWAIQ ACADEMY

استفساراتكم؟ 🤔