LANGUAGE	MODULE	REVISION	DRAWING NO.
العربية	علوم الحاسوب والبيانات	A	RS-003

بحث أصيل من بيانات عامة: دليل تحديد النطاق البحثي

كيف تطرح سؤالًا بحثيًا أصيلًا وتجيب عنه باستخدام مجموعات بيانات موجودة فعلًا — دون معدات، دون مختبر، دون تكلفة. الفئة الأكثر سهولة في الوصول إليها في هذا الدليل، والأسهل في إساءة تنفيذها دون معرفة المزالق.

المستوى: مبتدئ إلى متوسط | المدة الزمنية: ٣–٦ أسابيع | التكلفة: $0

01 الهدف

كل عام، تبقى آلاف مجموعات البيانات العامة على الإنترنت — إحصاءات حكومية، مستودعات علمية، بيانات صحية ومناخية مفتوحة — مجانية تمامًا ولم يلمسها تحليل أصيل. البحث الحاسوبي وبحث البيانات يعني طرح سؤال حقيقي على تلك البيانات والالتزام بالصدق والدقة حول ما تُظهره الإجابة فعليًا. إنها الفئة الأكثر سهولة في الوصول إليها في هذه السلسلة: تحتاج فقط حاسوبًا واتصالًا بالإنترنت، لا شيء آخر. وهي أيضًا الأسهل في الخطأ بطريقة محددة — تحويل "وصفتُ مجموعة بيانات" إلى "أجريتُ بحثًا أصيلًا" هو بالضبط ما يعلّمه هذا الدليل.

02 ما الذي يجعل المشروع قويًا في هذه الفئة

المشروع القوي يطرح سؤالًا لم تُبنَ مجموعة البيانات للإجابة عنه مباشرة، ويقوم بعمل أصيل للإجابة عنه — بدمج مجموعتي بيانات، أو حساب متغير مشتق جديد، أو اختبار فرضية محددة لم يختبرها ناشرو البيانات الأصليون أبدًا. أما المشروع الضعيف فيكتفي بإعادة وصف ما هو موجود بالفعل في البيانات ("هذا الرسم يوضح انبعاثات CO2 حسب الدولة" ملخص، وليس بحثًا — يمكن لأي شخص استخراج ذلك الرقم).

03 تحديد نطاق مشروعك

مثال تطبيقي — تضييق نطاق الموضوع

واسع جدًا: "تحليل بيانات المناخ العالمية."

لا يزال وصفًا فقط: "إظهار كيف تغيّرت درجات حرارة المدن خلال ٢٠ عامًا."

قابل للبحث: "هل هناك علاقة ارتباطية قابلة للقياس بين نسبة الغطاء الشجري في مدينة ومعدل دخول المستشفيات المرتبط بالحرارة فيها، بدمج مجموعة بيانات حراجة حضرية مفتوحة مع سجلات صحية عامة عبر ١٥ مدينة؟"

الصيغة القابلة للبحث تدمج مجموعتي بيانات مستقلتين لاختبار علاقة محددة — هذا الدمج هو حيث تكمن المساهمة الأصيلة.

04 المنهجية وجمع البيانات

- **إيجاد مجموعات بيانات موثوقة.** الوكالات الإحصائية الحكومية، وبيانات البنك الدولي المفتوحة، ومستودعات بيانات منظمة الصحة العالمية، ومكاتب التعداد الوطنية، والمستودعات العلمية الجامعية، جميعها مصادر قوية. يمكن أن يكون Kaggle مفيدًا لكن تحقق من المصدر الأصلي ورخصة أي مجموعة بيانات مستضافة هناك — ليس كل ما يُرفع منسوبًا بشكل صحيح.
- **تنظيف البيانات.** مجموعات البيانات الحقيقية تحتوي قيمًا مفقودة وتنسيقًا غير متسق وقيمًا شاذة. وثّق كل قرار تنظيف تتخذه (حذف صف، ملء فجوة) — سيسألك المحكّمون، و"حذفت الغريبة منها فقط" ليست إجابة يمكن الدفاع عنها.
- **الأدوات.** برامج جداول البيانات كافية للعديد من المشاريع؛ Python (مع pandas) أو R معياريان لأي شيء أكثر تقدمًا. أدوات مجانية تعمل بالمتصفح مثل Google Colab تزيل الحاجة لتثبيت أي شيء.

05 الامتثال والأخلاقيات — اقرأ قبل نشر أي نتائج

القاعدة التي تُوقع أغلب المشاريع البحثية للمبتدئين

يجب أن يصف ملخصك ولوحتك **تحليلك الأصيل فقط** — وليس مجموعة البيانات نفسها. يجب توثيق كل مجموعة بيانات تستخدمها بمصدرها ورخصتها بشكل صحيح، وإذا تطلبت مجموعة بيانات نسبًا محددًا أو اتفاقية استخدام، فأنت ملزم بها تمامًا كما لو كنت تستخدم كود أو كتابة شخص آخر.

- **وثّق كل مجموعة بيانات** بمصدرها المؤسسي وتاريخ الوصول إليها ورابطها — تعامل مع هذا بنفس جدية توثيق ورقة بحثية.
- **تحقق من الترخيص قبل استخدام أي شيء مقيّد تجاريًا** — "مرئي للعامة" لا يعني دائمًا "حر الاستخدام في مسابقة"، خاصة لبعض مجموعات البيانات المستضافة على Kaggle أو المُستخرجة آليًا.
- **كن واضحًا حول ما هو ملكك.** إذا استخدمت مكتبة كود لشخص آخر أو نهج درس تعليمي، وثّقه وكن واضحًا حول ما بنيته أنت مقابل ما اقتبسته — وصف عمل شخص آخر غير الموثّق كعمل أصيل في ملخص ISEF مشكلة نزاهة جدية، وليست أمرًا بسيطًا.

06 التحليل والتفسير

الفخ الشائع	كيف يبدو	الحل
عرض الارتباط كسببية	"X يسبب Y" من ارتباط ملاحظ فقط	اذكره كارتباط وناقش صراحة ما قد يفسره غير ذلك
الصيد الإحصائي / الانتقاء	اختبار متغيرات كثيرة حتى يبدو أحدها ذا دلالة، ثم الإبلاغ عنه فقط	حدد فرضيتك ومتغيراتك قبل إجراء التحليل، لا بعد رؤية النتائج
تحليل واسع النطاق أكثر من اللازم	محاولة تحليل ١٥ متغيرًا عبر ٣ مجموعات بيانات دفعة واحدة	اختر علاقة واضحة واحدة واختبرها بعمق بدلًا من عدة علاقات بسطحية

07 الأخطاء الشائعة

الخطأ	لماذا يمثل مشكلة	الحل
وصف مجموعة بيانات بدلًا من تحليلها	ليس بحثًا أصيلًا — سيلاحظ المحكّمون ذلك فورًا	اطرح دائمًا سؤالًا لا تجيب عنه البيانات الخام مباشرة
استخدام بيانات غير موثّقة أو غير مرخصة	مشكلة نزاهة، قد تُستبعد المشروع	وثّق كل مصدر، تحقق من كل رخصة قبل الاستخدام
عدم توثيق قرارات التنظيف	يجعل نتائجك مستحيلة التحقق أو الثقة بها	احتفظ بسجل مستمر لكل خطوة تنظيف بيانات

08 الخطوة التالية

من دليل تحديد النطاق إلى التقديم

التحليل الأصيل والدقيق للبيانات العامة يناسب فئتي ISEF الأحياء الحاسوبية والمعلوماتية الحيوية أو برمجيات الأنظمة، حسب موضوعك. راجع مخطط المشروع في دليل ISEF والجامعات لتحويل نتائجك إلى ملخص ولوحة متوافقين.

هذه الفئة أيضًا الجسر الأكثر طبيعية نحو فرع **أولمبياد البرمجة والمعلوماتية** المخطط له في MMI — مهارات التحليل هنا تتداخل مباشرة مع التفكير الخوارزمي.