توثيق مشروع تحليل بيانات

دوري روشن السعودي — موسم 2025

تحليل الحضور الجماهيري والكفاءة التنظيمية عبر Power Query وDAX

إعداد: محمد خالد

محلل بيانات

يوليو 2026

جدول المحتويات

الملخص التنفيذي 3

الفصل الأول: نظرة عامة على المشروع 4

الفصل الثاني: منهجية العمل 5

الفصل الثالث: تقرير فحص جودة البيانات 6

الفصل الرابع: معجم البيانات (Data Dictionary) 7

الفصل الخامس: نموذج البيانات المقترح 9

الفصل السادس: مؤشرات الأداء الرئيسية المقترحة (KPIs) 10

الفصل السابع: تصميم لوحة المعلومات المقترح 11

الفصل الثامن: التحديات وحدود التحليل 12

الفصل التاسع: التوصيات والخطوات القادمة 13

الخاتمة 13

الملحق: قائمة المصطلحات (Glossary) 14

الملخص التنفيذي

يقدّم هذا المستند التوثيق الكامل لمشروع تحليل وتصور بيانات دوري روشن السعودي لموسم 2025، بدءًا من تحديد أسئلة العمل، مرورًا بفحص وتنظيف البيانات الخام عبر Power Query، وصولًا إلى بناء نموذج بيانات موحّد يخدم لوحة معلومات تفاعلية تدعم قرارات إدارة الموارد اللوجستية في الملاعب.

تعتمد المنهجية المتبعة على دمج مصدرين خام (بيانات الحضور الجماهيري وبيانات المنظمين) في نموذج بيانات مسطّح (Flat Table) بعد معالجة أربع مشكلات جودة رئيسية، مما يتيح تحليلًا دقيقًا للعلاقة بين حجم الجمهور وحجم الموارد البشرية المخصصة لكل مباراة.

الفصل الأول: نظرة عامة على المشروع

1.1 مقدمة المشروع

يهدف هذا المشروع إلى دراسة وتحليل بيانات الحضور الجماهيري وبيانات أعداد المنظمين لمباريات دوري روشن السعودي لموسم 2025. تم جلب البيانات من «منصة البيانات السعودية المفتوحة» على هيئة ملفات خام (Raw Data) بصيغة إكسل، تغطي جميع الجولات والمباريات والملاعب المستضيفة للدوري.

1.2 نطاق المشروع ومصادر البيانات

- مصدر البيانات: منصة البيانات السعودية المفتوحة (Open Data Platform).
- الفترة الزمنية: موسم دوري روشن السعودي 2025 بكامل جولاته.
- الملفات الخام: ملف الحضور الجماهيري، وملف بيانات المنظمين، بصيغة Excel.
- أدوات المعالجة والتحليل: Power Query لتنظيف البيانات، ونموذج بيانات مبني على DAX لعرضها في لوحة معلومات تفاعلية.

1.3 أهداف المشروع

1. دمج البيانات: ربط ملف الحضور الجماهيري بملف المنظمين للخروج برؤية موحّدة وشاملة لكل مباراة.
2. تحليل الكفاءة اللوجستية: دراسة العلاقة بين حجم الجمهور وعدد المنظمين في كل ملعب لتقييم كفاءة التوزيع.
3. دعم صناع القرار: تحديد الجولات والملاعب الأكثر إقبالًا لتسهيل تخطيط وتوزيع الموارد البشرية مستقبلًا.

1.4 أسئلة العمل (Business Questions)

يسعى هذا التحليل إلى الإجابة على الأسئلة التالية عبر لوحة المؤشرات (Dashboard):

4. ما هي الجولة الأعلى حضورًا جماهيريًا، وهل هي نفسها الأعلى من حيث عدد المنظمين؟
5. ما هو الملعب الأكثر كفاءة (أعلى حضور بأقل عدد منظمين)؟
6. من هو الفريق الذي تشهد مبارياته أعلى إقبال جماهيري؟
7. هل توجد علاقة طردية ثابتة بين زيادة الجمهور وزيادة عدد المنظمين؟

1.5 أصحاب المصلحة المستفيدون

يخدم هذا التحليل عدة جهات معنية بتنظيم وإدارة مباريات الدوري:

- الاتحاد السعودي لكرة القدم — لتقييم أداء الجولات والملاعب على مستوى الدوري.
- إدارات الملاعب والمنشآت الرياضية — لتخطيط الموارد اللوجستية والأمنية لكل مباراة.
- فرق التسويق والرعاية التجارية — لفهم أنماط الإقبال الجماهيري لكل نادٍ.
- محللو البيانات وصنّاع القرار — للاستفادة من لوحة المعلومات في التخطيط المستقبلي.

الفصل الثاني: منهجية العمل

2.1 أدوات وتقنيات التحليل

الاستخدام في المشروع	الأداة
تنظيف البيانات، فك دمج الخلايا، تقسيم الأعمدة، ودمج الاستعلامات	Power Query
بناء العلاقات بين الجداول عبر معرّف المباراة (ID_Match)	نموذج البيانات (Data Model)
كتابة المقاييس المحسوبة (Measures) الخاصة بمؤشرات الأداء	DAX
عرض التحليلات بصريًا ودعم اتخاذ القرار التفاعلي	لوحة المعلومات (Dashboard)

2.2 مراحل العمل (منهجية ETL)

اتُّبعت في هذا المشروع منهجية استخراج وتحويل وتحميل البيانات (Extract – Transform – Load) عبر المراحل التالية:

8. الاستخراج (Extract): سحب الملفات الخام من منصة البيانات المفتوحة بصيغة Excel.
9. الفحص (Assess): فحص بصري ومنطقي لاكتشاف مشكلات الجودة والتنسيق في كل ملف.
10. التحويل (Transform): تنظيف البيانات عبر Power Query (إزالة الترويسات، فك الدمج، تقسيم الأعمدة، التحقق من الأنواع).
11. الدمج (Merge): ربط جدولي الحضور والمنظمين في نموذج بيانات موحّد عبر معرّف المباراة.
12. التحميل (Load): تحميل البيانات النظيفة إلى نموذج البيانات النهائي المستخدم في لوحة المعلومات.
13. التحليل والعرض (Analyze & Visualize): كتابة مقاييس DAX وتصميم لوحة المعلومات التفاعلية.

2.3 معايير قبول الجودة (Definition of Done)

تُعتبر عملية التنظيف مكتملة عندما تتحقق المعايير التالية في نموذج البيانات النهائي:

- عدم وجود صفوف إجماليات فرعية أو ترويسات وصفية ضمن جدول البيانات.
- عدم وجود قيم فارغة (Null) في عمود الجولة أو أي عمود مفتاحي.
- استقلالية كل عمود تحليلي (فريق مستضيف، فريق ضيف، مدينة، ملعب) في حقل منفصل.
- تطابق عدد المباريات في جدولي الحضور والمنظمين بعد الدمج دون فقدان أو تكرار سجلات.

الفصل الثالث: تقرير فحص جودة البيانات

بناءً على الفحص البصري والمنطقي للملفات الخام لدوري روشن 2025، تم رصد أربع مشكلات جودة رئيسية، وتحديد الإجراء التصحيحي لكل منها ضمن بيئة Power Query، مع تصنيف درجة أثر كل مشكلة على دقة التحليل النهائي:

1 مشكلة الأسطر العلوية والإجماليات الفرعية

درجة الأثر: مرتفع

الوصف: يحتوي الملف على ترويسة علوية (شعار وعنوان) وصفية، وأسطر إجماليات فرعية مدمجة تفصل بين الجولات.

الإجراء في Power Query: إزالة الأسطر العلوية الزائدة، وتصفية (Filter) الجدول لاستبعاد أسطر المجاميع الفرعية.

الهدف: ضمان احتواء الجدول على بيانات المباريات الفعلية فقط، ومنع تضاعف الأرقام (Double Counting) عند الحساب.

2 مشكلة الخلايا المدمجة في عمود «الجولة»

درجة الأثر: مرتفع

الوصف: اسم الجولة مكتوب مرة واحدة ومدمج على مستوى 9 مباريات، مما يترك بقية الصفوف فارغة للآلة.

الإجراء في Power Query: فك الدمج وتطبيق أمر الملء للأسفل (Fill Down) ليتكرر اسم الجولة في كل الأسطر المقابلة لها.

الهدف: تمكين المستخدم من تصفية لوحة المعلومات بناءً على الجولة بشكل صحيح دون سقوط أي مباراة من الحسابات.

3 مشكلة دمج النصوص (عمود الفريقين وعمود المنشأة)

درجة الأثر: متوسط

الوصف: عمود «الفريقين» يدمج (المستضيف × الضيف)، وعمود «المنشأة» يدمج اسم الملعب مع المدينة بين قوسين.

الإجراء في Power Query: استخدام خاصية تقسيم العمود (Split Column) بناءً على علامة × للفرق، وبناءً على الأقواس للملاعب.

الهدف: استخراج (الفريق المستضيف، الفريق الضيف، المدينة، الملعب) كأعمدة مستقلة، لإتاحة مقارنات متقدمة كأداء النادي على أرضه مقابل خارج أرضه.

4 مشكلة دمج الملفين المنفصلين

درجة الأثر: مرتفع

الوصف: بيانات الحضور في ملف، وبيانات المنظمين في ملف آخر، والمستخدم يحتاج رؤيتهما معًا.

الإجراء في Power Query: دمج الاستعلامات (Merge Queries) بناءً على رقم المباراة (م) أو توليفة (التاريخ + الفرق) لجمع الحضور والمنظمين في سطر واحد لكل مباراة.

الهدف: توحيد مصادر البيانات لبناء نموذج مبسّط وخفيف (Flat Table أو Star Schema) يسهّل كتابة معادلات DAX وتصميم الرسوم البيانية.

الفصل الرابع: معجم البيانات (Data Dictionary)

4.1 جدول الحضور الجماهيري

الوصف والهدف	نوع البيانات المتوقع	الاسم المقترح بالإنجليزية	اسم العمود الأصلي
الرقم التسلسلي للمباراة؛ يُستخدم لربط هذا الجدول بجدول المنظمين مستقبلًا.	رقم صحيح (Integer)	ID_Match	م
اسم الملعب والمدينة المقامة فيها المباراة؛ يُستخدم لتصنيف الحضور حسب الملاعب.	نص (Text)	Name_Stadium	اسم المنشأة
تاريخ يوم المباراة؛ يمكّن من التحليل الزمني حسب الأيام والأشهر.	تاريخ (Date)	Date_Match	التاريخ
رقم الجولة في الدوري (الأولى، الثانية...)؛ يُستخدم لقياس تغيّر الحضور مع تقدّم الدوري.	نص (Text)	Name_Round	الجولة
اسم الناديين المتبارين؛ سيُقسّم لاحقًا لمعرفة شعبية الحضور لكل نادٍ بشكل مستقل.	نص (Text)	Teams	الفريقين
العدد الفعلي للمشجعين في الملعب؛ الحقل الأساسي لحساب إجمالي ومتوسط الحضور الجماهيري.	رقم صحيح (Integer)	Count_Attendance	الحضور الجماهيري

4.2 جدول المنظمين

الوصف والهدف	نوع البيانات المتوقع	الاسم المقترح بالإنجليزية	اسم العمود الأصلي
الرقم التسلسلي للمباراة؛ يُستخدم لربط هذا الجدول بجدول الحضور.	رقم صحيح (Integer)	ID_Match	م
اسم الملعب والمدينة المقامة فيها المباراة؛ يُستخدم لتصنيف البيانات حسب الملاعب.	نص (Text)	Name_Stadium	اسم المنشأة
تاريخ يوم المباراة؛ يمكّن من التحليل الزمني حسب الأيام والأشهر.	تاريخ (Date)	Date_Match	التاريخ
رقم الجولة في الدوري (الأولى، الثانية...)؛ يُستخدم لقياس تغيّر الحضور مع تقدّم الدوري.	نص (Text)	Round_Name	الجولة
اسم الناديين المتبارين؛ سيُقسّم لاحقًا لمعرفة شعبية الحضور لكل نادٍ بشكل مستقل.	نص (Text)	Teams	الفريقين
العدد الفعلي للمنظمين المسؤولين عن إدارة المباراة؛ الحقل الأساسي لحساب الموارد البشرية ومقارنتها بالحضور.	رقم صحيح (Integer)	Organizers_Count	عدد المنظمين

ملاحظة على الأعمدة المحذوفة

تم إهمال أعمدة (المسابقة أو الفعالية) و(إجمالي الحضور / المنظمين الفرعي) من المعجم؛ نظرًا لاتخاذ قرار مسبق بحذفها في مرحلة التنظيف لعدم جدواها في بناء نموذج البيانات.

الفصل الخامس: نموذج البيانات المقترح

5.1 العلاقة بين الجدولين

يقوم نموذج البيانات على ربط جدول «الحضور الجماهيري» بجدول «المنظمين» عبر مفتاح مشترك هو معرّف المباراة (ID_Match)، بعلاقة من نوع واحد إلى واحد (One-to-One)، نظرًا لأن كل مباراة تُمثَّل بسجل واحد في كل جدول.

الجدول الهدف	نوع العلاقة	الحقل المفتاحي	الجدول المصدر
جدول المنظمين	واحد إلى واحد (1:1)	ID_Match	جدول الحضور الجماهيري

5.2 مخرجات الدمج: الجدول الموحّد (Flat Table)

بعد الدمج، يحتوي الجدول الموحّد على الحقول التالية والتي تشكّل الأساس لجميع مقاييس DAX ومرئيات لوحة المعلومات:

- معرّفات المباراة والتاريخ والجولة.
- الفريق المستضيف والفريق الضيف (بعد تقسيم عمود الفريقين).
- الملعب والمدينة (بعد تقسيم عمود المنشأة).
- عدد الحضور الجماهيري.
- عدد المنظمين.
- حقل محسوب: نسبة الكفاءة التنظيمية = الحضور الجماهيري ÷ عدد المنظمين.

الفصل السادس: مؤشرات الأداء الرئيسية المقترحة (KPIs)

لضمان أن تجيب لوحة المعلومات بشكل مباشر على أسئلة العمل الأربعة، يُقترح بناء مقاييس DAX التالية ضمن نموذج البيانات:

الغرض التحليلي	طريقة الحساب المقترحة	المؤشر
قياس الحجم الكلي للجمهور طوال الموسم.	SUM(Count_Attendance)	إجمالي الحضور الجماهيري
تقييم مستوى الإقبال العام بشكل نمطي.	AVERAGE(Count_Attendance)	متوسط الحضور لكل مباراة
قياس الحجم الكلي للموارد البشرية المستخدمة.	SUM(Organizers_Count)	إجمالي عدد المنظمين
تحديد الملاعب الأكثر كفاءة في استغلال الموارد.	الحضور ÷ عدد المنظمين	نسبة الكفاءة التنظيمية
الإجابة على السؤال الأول من أسئلة العمل.	MAX عبر Name_Round	الجولة الأعلى حضورًا
الإجابة على السؤال الثالث من أسئلة العمل.	متوسط الحضور حسب الفريق المستضيف	الفريق الأعلى استقطابًا للجمهور
الإجابة على السؤال الرابع؛ قياس قوة العلاقة الطردية.	CORREL أو دالة ارتباط DAX مخصصة	معامل الارتباط (الحضور × المنظمين)

الفصل السابع: تصميم لوحة المعلومات المقترح

7.1 عناصر اللوحة الرئيسية

14. بطاقات مؤشرات (KPI Cards) أعلى اللوحة: إجمالي الحضور، إجمالي المنظمين، متوسط نسبة الكفاءة، عدد الجولات المكتملة.
15. رسم بياني عمودي (Bar Chart) يوضح الحضور الجماهيري لكل جولة، مع تراكب خط لعدد المنظمين لتسهيل المقارنة البصرية.
16. رسم مبعثر (Scatter Chart) يوضح العلاقة بين الحضور وعدد المنظمين لكل مباراة، لدعم الإجابة على سؤال العلاقة الطردية.
17. جدول أو رسم شريطي (Bar/Table) يرتّب الملاعب حسب نسبة الكفاءة التنظيمية من الأعلى إلى الأقل.
18. رسم بياني لأعلى الفرق استقطابًا للجمهور، مع إمكانية التصفية حسب كون الفريق مستضيفًا أو ضيفًا.
19. عوامل تصفية (Slicers) تفاعلية: الجولة، الملعب، المدينة، والفريق.

7.2 مبادئ تصميم مقترحة

- استخدام هوية لونية موحّدة مستوحاة من ألوان دوري روشن السعودي لتعزيز الطابع الاحترافي للوحة.
- ترتيب العناصر من الأهم إلى الأقل أهمية بما يتوافق مع أسئلة العمل الأربعة بشكل مباشر.
- إضافة تلميحات (Tooltips) توضيحية على كل مؤشر لتسهيل قراءة البيانات من قبل غير المختصين.
- دعم التصدير (Export) لتقارير PDF أو PowerPoint لعرضها على صنّاع القرار بشكل دوري.

الفصل الثامن: التحديات وحدود التحليل

كأي مشروع تحليل بيانات يعتمد على مصادر مفتوحة، توجد بعض المحاذير التي ينبغي مراعاتها عند تفسير نتائج هذا التحليل:

إجراء التخفيف المقترح	درجة الأثر	الوصف	نوع المخاطرة
مطابقة عدد السجلات بين الجدولين بعد الدمج والتحقق من الفروقات	متوسط	احتمال وجود مباريات ناقصة أو غير مسجّلة في أحد الملفين	اكتمال البيانات
توضيح مصدر الرقم في توثيق اللوحة، وعدم استخدامه كمقياس دقيق للحضور الفعلي	متوسط	الأرقام المعلنة قد تمثل التذاكر المباعة وليس الحضور الفعلي داخل الملعب	دقة الحضور المُعلن
عرض المتوسطات المتحركة إلى جانب الأرقام المطلقة	منخفض	تركّز بعض المباريات الكبرى في جولات معينة قد يشوّه المقارنات بين الجولات	تحيّز موسمي
تمييز القيم الشاذة بصريًا في لوحة المعلومات وعدم حذفها	متوسط	مباريات القمة قد تُظهر حضورًا استثنائيًا يؤثر على متوسط التحليل العام	القيم الشاذة (Outliers)

الفصل التاسع: التوصيات والخطوات القادمة

20. اعتماد نموذج البيانات الموحّد (Flat Table) كمصدر رسمي وحيد للحقيقة (Single Source of Truth) لجميع تقارير الحضور والتنظيم.
21. أتمتة عملية التحديث (Refresh) الدورية للبيانات عند صدور بيانات جولات جديدة من المنصة المفتوحة.
22. توسيع نطاق التحليل مستقبلًا ليشمل بيانات إضافية مثل الحوادث الأمنية أو زمن دخول الجمهور، لتعميق تحليل الكفاءة اللوجستية.
23. ربط لوحة المعلومات بمؤشرات الأداء الخاصة بالأندية (كالنتائج والترتيب) لدراسة العلاقة بين الأداء الرياضي والإقبال الجماهيري.
24. إعداد تقرير دوري (أسبوعي أو بعد كل جولة) يُرسل تلقائيًا لصنّاع القرار يلخّص أهم المؤشرات والتنبيهات.

الخاتمة

يمثّل هذا المشروع نموذجًا متكاملًا لدورة حياة تحليل البيانات، بدءًا من تحديد أسئلة العمل، مرورًا بفحص وتنظيف البيانات الخام ومعالجة مشكلاتها الأربع الرئيسية عبر Power Query، وصولًا إلى بناء نموذج بيانات موحّد ومؤشرات أداء رئيسية تخدم لوحة معلومات تفاعلية. ويوفّر هذا التوثيق أساسًا صلبًا يمكن البناء عليه لتوسيع نطاق التحليل وتحسين كفاءة إدارة الموارد اللوجستية لمباريات دوري روشن السعودي في المواسم القادمة.

الملحق: قائمة المصطلحات (Glossary)

المصطلح	التعريف
ETL	Extract, Transform, Load — دورة استخراج البيانات وتحويلها وتحميلها في نموذج التحليل.
Power Query	أداة ضمن Excel وPower BI لتنظيف البيانات وتحويلها قبل التحليل.
DAX	لغة صيغ تحليلية (Data Analysis Expressions) تُستخدم لبناء المقاييس المحسوبة في نماذج البيانات.
Flat Table	جدول بيانات موحّد ناتج عن دمج عدة مصادر في هيكل واحد مبسّط.
Star Schema	نموذج بيانات يتكوّن من جدول حقائق مركزي مرتبط بجداول أبعاد وصفية.
KPI	Key Performance Indicator — مؤشر أداء رئيسي يُستخدم لقياس مدى تحقق الأهداف.
Fill Down	أمر في Power Query يكرّر آخر قيمة معروفة في الخلايا الفارغة أسفلها.
Merge Queries	عملية دمج جدولين أو أكثر في Power Query بناءً على حقل مشترك.