



مشروع تخرج NEXT22 \_ DKH1\_DAT2\_G1E

#### • أعضاء الفريق: -

- 1. أسعد السعدني بدير عبد الكريم
  - 2. رضا إبراهيم الزكي محمد
  - 3. أحمد فتحي محمود العدل
  - 4. أمنية عبد العزيز إبراهيم
    - 5. غاده خضر أبو بكر
    - 6. اميره اشرف الجمل





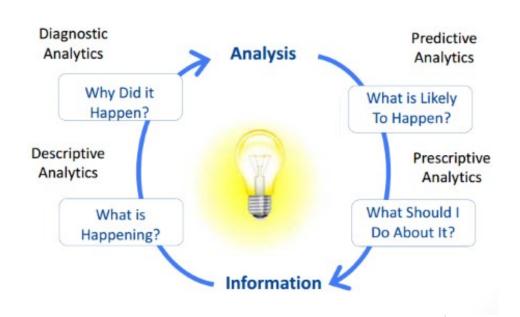
### تحليل البيانات

• هو عملية منهجية تجمع بين جمع البيانات وتنظيمها وتحويلها وتحليلها بهدف استخلاص معلومات قيمة واتخاذ قرارات مدروسة. بعبارة أبسط، هو عملية استخراج المعنى من الأرقام والحقائق.

### أهمية تحليل البيانات

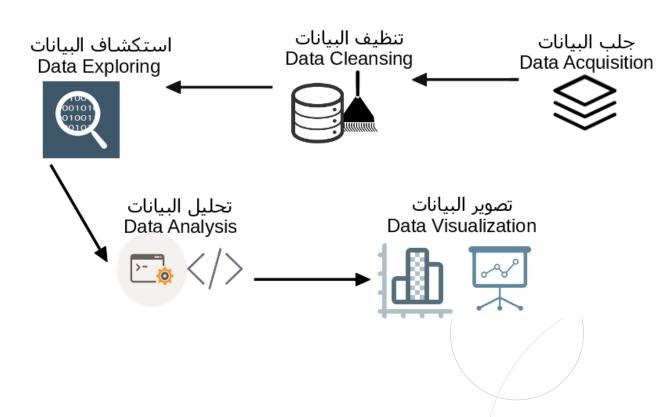
- اتخاذ قرارات أفضل: يوفر تحليل البيانات رؤى عميقة حول البيانات، مما يساعد في اتخاذ قرارات أكثر استنارة وفعالية.
  - اكتشاف الفرص: يساعد في تحديد الاتجاهات والأنماط الخفية في البيانات، مما يفتح آفاقًا جديدة للنمو والتطوير.
- تحسين العمليات: يمكن تحليل البيانات لتحديد نقاط الضعف والتحسين في العمليات الحالية، مما يؤدي إلى زيادة الكفاءة والفعالية.
  - تخصيص الخدمات: يمكن استخدام تحليل البيانات لتقديم خدمات مخصصة للعملاء، مما يعزز رضا العملاء وولائهم.
  - الابتكار: يمكن لتحليل البيانات أن يشعل شرارة الابتكار من خلال تحديد فرص جديدة وتطوير منتجات وخدمات جديدة.

### أنواع تحليل البيانات



- هناك العديد من أنواع تحليل البيانات، ولكل نوع هدف وطريقة تحليل مختلفة. إليك بعض الأنواع الشائعة:
  - التحليل الوصفي: يهدف إلى وصف البيانات وتلخيصها، مثل حساب المتوسط والانحراف المعياري.
- التحليل التشخيصي: يهدف إلى فهم سبب حدوث شيء ما، مثل تحديد العوامل التي تؤثر على المبيعات.
  - التحليل التنبؤي: يهدف إلى التنبؤ بالأحداث المستقبلية، مثل التنبؤ بالمبيعات في المستقبل.
  - التحليل التكيفي: يهدف إلى تحسين النماذج والتنبؤات باستمرار مع مرور الوقت.

### خطوات تحليل البيانات



- •جمع البيانات "يتم جمع البيانات من مصادر مختلفة، مثل قواعد البيانات وجداول البيانات وملفات النصوص .
  - تنظيف البيانات ! يتم تنظيف البيانات لإزالة الأخطاء والقيم المفقودة والمتناقضة .
  - •تحويل البيانات ليتم تحويل البيانات إلى تنسيق مناسب للتحليل .
- •استكشاف البيانات "يتم استكشاف البيانات للكشف عن الأنماط والاتجاهات .
  - •نمذجة البيانات ليتم بناء نماذج إحصائية أو تعلم آلي لتحليل البيانات .
    - •تقييم النتائج ليتم تقييم النتائج للتأكد من دقتها وموثوقيتها .
    - •تقديم النتائج ليتم تقديم النتائج في شكل تقارير أو رسوم بيانية .

### أدوات تحليل البيانات

- 1. أدوات التحليل الإحصائي:
- تستخدم هذه الأدوات لإجراء التحليلات الإحصائية المعقدة على البيانات، مثل اختبار الفرضيات، والتحليل الانحداري، وتحليل التباين.
  - برامج: Stata. ، R ، SPSS ، SAS
  - مكتبات برمجة: Statsmodels ، SciPy في Statsmodels ،
    - 2أدوات التصور:
  - تتيح هذه الأدوات تحويل البيانات الرقمية إلى رسوم بيانية وتفاعلية، مما يسهل فهم الأنماط والاتجاهات.
    - برامج: Tableau, Power BI, Qlik Sense, Google Data Studio.
      - مكتبات برمجة: Matplotlib, Seaborn, Plotly في Matplotlib.
        - .3أدوات التعلم الآلي:
  - تستخدم هذه الأدوات لبناء نماذج تنبؤية وتصنيفية، مثل نماذج الانحدار اللوجستي، والشبكات العصبية، وتحليل المكونات الرئيسية.
    - برامج: .TensorFlow, PyTorch, Scikit-learn
      - لغات برمجة: . Python, R

### أدوات تحليل البيانات

- 4. أدوات قواعد البيانات:
- تستخدم هذه الأدوات لإدارة وتخزين وتنظيم البيانات، وتوفير استعلامات للوصول إليها.
  - برامج: MySQL, PostgreSQL, SQL Server, MongoDB.
    - لغات استعلام: .SQL
    - .5أدوات تحليل البيانات الضخمة:
- تستخدم هذه الأدوات لتحليل كميات هائلة من البيانات، وتتطلب عادةً موارد حوسبة كبيرة.
  - برامج: . Apache Hadoop, Apache Spark, Apache Flink
    - لغات برمجة: . Scala, Python, Java
      - 6أدوات التحليل النصي:
- تستخدم هذه الأدوات لتحليل البيانات النصية، مثل تحليل المشاعر، وتصنيف النصوص، واستخراج الكلمات المفتاحية.
  - مكتبات برمجة: NLTK, spaCy في NLTK, spaCy
    - 7أدوات التحليل الجغرافي:
  - تستخدم هذه الأدوات لتحليل البيانات المرتبطة بالموقع الجغرافي، مثل البيانات الديموغرافية، وتحليل النمط المكاني.
    - برامج: .ArcGIS, QGIS

### أدوات تحليل البيانات

- عوامل اختيار الأداة المناسبة:
- نوع البیانات: هل هی بیانات رقمیة، نصیة، جغرافیة، أم مزیج منها؟
  - حجم البيانات: هل هي بيانات صغيرة، متوسطة، أم ضخمة؟
- نوع التحليل المطلوب: هل تريد إجراء تحليل إحصائي، أو تصور البيانات، أو بناء نموذج تعلم آلي؟
  - المهارات التقنية للمستخدم: ما هي لغات البرمجة والمكتبات التي يجيدها المستخدم؟
    - الميزانية المتاحة: ما هي تكلفة الأدوات والبرامج؟
- ملاحظة: التكامل بين الأدوات: غالبًا ما يتم استخدام عدة أدوات معًا لإكمال عملية تحليل البيانات. على سبيل المثال، يمكن استخدام قاعدة بيانات لتخزين البيانات، ثم استخدام Power-Bi لتصور النتائج.



باور بي آي هي أداة تحليلية قوية ومرنة من مايكروسوفت، صُممت لتسهيل عملية تحويل البيانات الخام إلى رؤى قابلة للعمل. تتميز بواجهتها البسيطة وقدرتها على إنشاء لوحات تحكم تفاعلية جذابة، مما يجعلها أداة شائعة الاستخدام في مختلف الصناعات والشركات.





### باور بي آي

#### مكان باور بي آي من أدوات التحليل

• باور بي آي من مايكروسوفت قد حجزت لنفسها مكانة مميزة في عالم أدوات تحليل البيانات.

#### لماذا باور بي آي؟

- •سهلة الاستخدام للا تحتاج لخبرة برمجية .
  - •مرنة :تتكامل مع مصادر بيانات متنوعة .
- •بصرية تقدم رسوم بيانية تفاعلية .
  - •تنبؤية تساعد في اتخاذ قرارات
    - مستقبلية .
- •تعاونية :تسمح بمشاركة البيانات .
  - •اقتصادية ـ تتوفر نسخة مجانية .





### تحليل بيانات الصحة الإنجابية والأم والطفل في مد فظة الدقهلية

مشروع تخرج NEXT22 \_ DKH1\_DAT2\_G1E

#### • أعضاء الفريق: -

- 1. أسعد السعدني بدير عبد الكريم
  - 2. رضا إبراهيم الزكي محمد
  - 3. أحمد فتحي محمود العدل
  - 4. أمنية عبد العزيز إبراهيم
    - 5. غاده خضر أبو بكر
    - 6. ميره اشرف الجمل



#### مقدمة



تحليل البيانات الصحية، وخاصة تلك المتعلقة بصحة الأم والطفل، يعد أمرًا بالغ الأهمية في تطوير وتحسين الخدمات الصحية وصنع السياسات الفعالة. في سياق محافظة الدقهلية بمصر، يكتسب هذا التحليل أهمية خاصة نظرًا لتنوع المناطق الجغرافية والتفاوت في توزيع الموارد الصحية. إن فهم الاتجاهات والعلاقات بين مختلف العوامل الصحية يمكّن بمصر، يكتسب هذا التحليل أهمية خاصة نظرًا لتنوع المناطق الجغرافية والتفاوت في توزيع الموارد الصحية السكان.

### أهميته لصانع القرار

- تحديد الفجوات في الخدمات الصحية:
- تحديد المناطق التي تفتقر إلى الخدمات الصحية الأساسية، مما يتيح لصانعي القرار اتخاذ إجراءات تصحيحية
  - توجيه الموارد بشكل أكثر فعالية:
  - توجيه الموارد المحدودة إلى المناطق التي تحتاجها.
    - تصميم برامج التدخل المناسبة:
  - يسمح التحليل بفهم العوامل المؤثرة على صحة الأم والطفل، مما يتيح تصميم برامج تدخل فعالة.

#### أهميته لصانع القرار

- توفير معلومات موثوقة لصنع القرار:
- يقدم التحليل معلومات دقيقة وموثوقة تساعد صانعي القرار في اتخاذ قرارات مستنيرة
  - تعزيز التوعية الصحية:
- من خلال التحليل تحديد المناطق التي تحتاج إلى برامج توعية صحية مكثفة، مما يساهم في تحسين التوعية.
  - تحسين الاستجابة للأزمات الصحية:
- تحديد المناطق التي قد تكون أكثر عرضة للأزمات الصحية، مما يتيح لصانعي القرار اتخاذ إجراءات وقائية واستجابية فعالة.

### منهجية البحث

تم اتباع منهجية دقيقة وشاملة في هذا البحث لضمان دقة النتائج وموثوقيتها. فيما يلي الخطوات الرئيسية للمنهجية المتبعة:

جمع البيانات: تم الحصول على عشرة ملفات من مديرية الصحة بالدقهلية، تحتوي على بيانات متنوعة تغطي جوانب مختلفة من صحة الأم والطفل في المحافظة.

#### معالجة البيانات:

تمت ترجمة بيانات الحقول لضمان اتساق المصطلحات وسهولة التحليل. استُخدم Python لتجهيز البيانات وتجميعها من الملفات المختلفة، مما سمح بدمج المعلومات من مصادر متعددة في مجموعة بيانات موحدة وقابلة للتحليل.



# استخدام البايثون

```
1 [1]: def Mapper (data, k, v):
           dicts = []
           for d in data:
               limited = d[[k,v]]
               lst = limited.set_index(k).T.to_dict('list')
               dicts.append(lst)
           return dicts
[2]: def Reducer(dicts, specific_k):
           result = 0
           for d in dicts:
               specific_v = d[specific_k]
               result = result + int(specific_v[0])
           return result
1 [3]: import pandas as pd
       d2015 = pd.read_csv('2015.csv')
       d2016 = pd.read_csv('2016.csv')
       d2017 = pd.read_csv('2017.csv')
       d2018 = pd.read_csv('2018.csv')
```

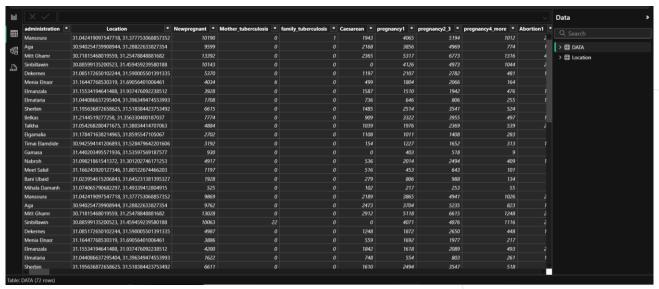
# ترجمة بيانات الأعمدة

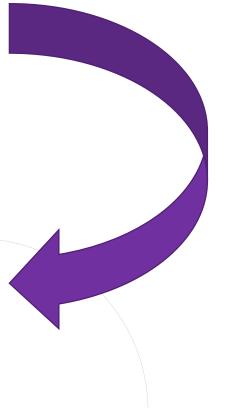
Management New pregnant Drain for the mother Drain for the family Caesarean first pregnancy Second or third pregnancy Fourth pregnancy and more Abortion once Abortion more than once Born alive The baby is dead Death of a child less than 7 days Blood pressure 140/90 and above Blood pressure is less than 140/90 Thyroid gland hypertrophy No Thyroid gland hypertrophy Heart cases Chest cases Polyethylene cases natural Hemoglobin less than 9 g Hemoglobin from 9 to less than 11 g Hemoglobin 11 g or more Negative Reese Positive Reese negative Wazerman positive Wazerman Negative sugar positive Sugar Negative albumin

الإدارة جملة حوامل جدد در ن للأم در ن للأسرة قيصرية حمل بكرية حمل ثاني أو ثالث حمل رابع فأكثر إجهاض مرة واحدة إجهاض أكثر من مرة مولود حي مو لو د میت وفيات المواليد في عمر أقل من ٧ أيام ضغط دم ۱٤٠ / ۹۰ فأكثر ضغط دم أقل من ١٤٠ / ٩٠ تضخم غدة درقية لا تضخم غدة در قية حالات قليبة حالات صدر بة حالات جهاز بولي هيمو جلو بين اقل من ٩جم هيمو جلوبين من ١٩لي اقل من ١١جم هیموجلوبین ۱ اجم فأكثر ريزس سلبي ريزس ايجابى وازرمان سلبي وازرمان ايجابي سكر سلبى سكر ايجابى ز لال سلبي

## من الإكسل إلى الباور بي آي

								l		l			l			1	6
ضخم غدة درقية		تضخم ع	ضغط دم		وفيات المواليد	مواليدسابقة		مرات الإجهاض		الحمل الحالى			تاريخ مرضى				7
	ч	نعم	ضغط دم أقل من 140 /	ضغط دم 140 / 90 فأكثر	فی عمرائل من7أیام	مولود ميت	مولود حی	أكثر من مرة	مرة واحدة	رابع فأكثر	ثانی أو ثالث	بكرية	قيصرية	درن للأسرة	درن للأم	جملة حوامل جدد	الادارة
_	11501	4	90 11496	69	38	13	7161	109	317	1097	5940	4534	2903	0	0	11584	8 9 المنصورة
_	10287	1	10268	20	29	54	6687	54	200	1019	5724	3545	3510	0	0	10288	ر <u>حصور</u> . 10 اجا
_	12792	116	12835	49	27	75	8208	183	425	1339	6900	4672	3437	0	0	12933	11 ميت غبر
_	10217	0	10358	20	13	26	6509	110	303	1176	5267	3942	0	0	0	10385	12 السنبلاوين
	4996	11	5070	14	11	20	4289	81	186	523	2645	1915	1486	1	0	5084	13 نکرنس
	3774	0	3768	1	1	3	2161	11	27	168	2004	1602	692	0	0	3774	14 منية النصر
	4143	69	4209	0	0	34	2571	81	162	561	2040	1611	1977	0	0	4212	15 المنزلة
	1479	0	1477	2	11	10	941	39	83	195	754	530	763	1	0	1479	16 المطرية
	6680	5	6660	24	6	35	4336	52	106	602	3767	2305	1930	0	0	6687	17 شربين
	8459	36	8477	19	29	51	5210	115	214	823	4444	3229	1815	2	0	8496	18 بلقاس
_	5156	2	5148	10	16	55	3235	112	216	649	2567	1941	1499	1	0	5158	19 طلخا
_	2813	22	2819	6	36	7	1816	46	101	377	1454	1016	1474	0	0	2847	20 الجمالية
_	3544	0	3543	1	18	15	2221	68	130	376	1892	1279	351	0	0	3543	21 تمى الأمديد
	1036	0	1052	0	0	1	708	0	2	50	614	395	49	0	0	1056	22 جىصة
_	4831	143	4941	33	7	21	2985	49	128	471	2576	1930	948	0	0	4977	23 نبروه
_	1162	1	1162	2	7	8	780	18	38	141	628	395	647	0	0	1164	24 میت سلسیل
_	1902	5	1906	1	1	1	1126	7	40	135	1040	732	356	0	0	1907	25 بنی عبید
_	807	0	802	5	1	2	486	9	14	50	436	321	263	0	1	807	26 مطة دمنة
_	95579	415	95991	276	251	431	61430	1144	2692	9752	50692	35894	24100	5	1	96381	27 الاجمالي







### منهجية البحث

تحليل البيانات: بعد تجهيز البيانات، تم نقلها إلى برنامج Power Bl لإنشاء لوحات المعلومات ( Dashboards) والإجابة على الأسئلة البحثية المحددة. هذا البرنامج سمح بإجراء تحليلات متقدمة وإنشاء تصورات بيانية تفاعلية.

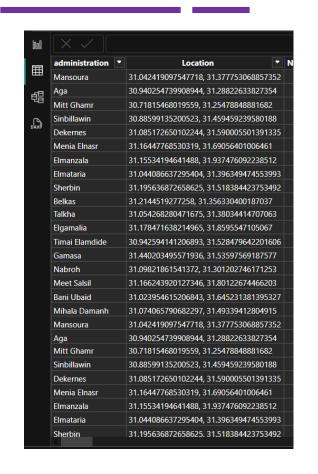
التحليل المكاني: تم ربط التحليلات بالمواقع الجغرافية للإدارات الصحية المختلفة. هذا الجانب من التحليل يعد حاسمًا لأنه يمكن صانعى القرار من:

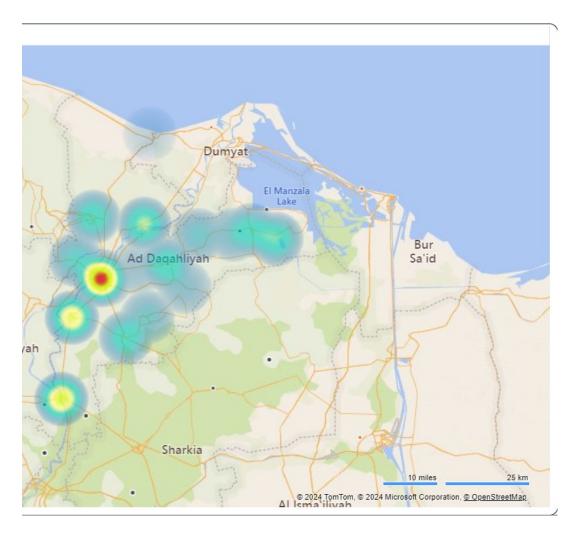
فهم التوزيع المكاني للإمكانات الطبية، بما في ذلك التطعيمات، والأطباء، والتمريض، وغيرها من الموارد الصحية.

تحديد المناطق التي تحتاج إلى تعزيز في الخدمات الصحية. توجيه جهود التوعية الصحية بشكل أكثر فعالية

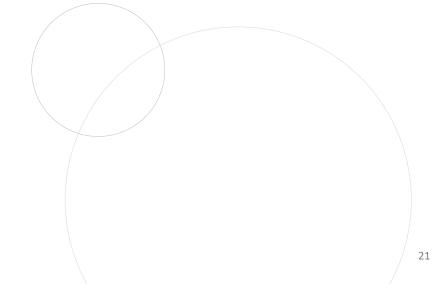


# ربط التحليلات بالمواقع الجغرافية





# أهم البيانات



### التاريخ المرضي



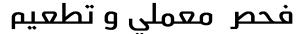
#### الأسئلة البحثية الرئيسية

- ما هي العلاقة بين الأمراض المزمنة (مثل السكري والضغط) ونتائج الحمل
  - ما هي نسبة السيدات الحوامل اللاتي يعانين من نقص الحديد؟
- توزیع الحمل علی مرات الحمل, مع الترکیز علی أربع مرات فأكثر و ربطها مكانیا بالتالي نعرف توزیع حملات التوعیة؟



#### البيانات المتوفرة

درن، قيصرية، حمل حالي، إجهاضات، مواليد سابقة، ضغط دم، تضخم الغدة الدرقية، هيموجلوبين





#### الأسئلة البحثية الرئيسية

- ما هي نسبة السيدات الحوامل المصابات بالأنيميا أو السكري
  - ما هي تغطية التطعيمات الروتينية؟
- ما هي العوامل المرتبطة بحالات الحمل الخطرة؟



زلال، سكر، وازرمان، ريزس، حالة التطعيم، حالات الحمل الخطرة، حديد+فوليك

### الحالة الصحية للسيدات الحوامل



#### الأسئلة البحثية الرئيسية

- ما هي العلاقة بين زيادة الوزن غير
   الطبيعية ومضاعفات الحمل
- ما هي نسبة السيدات الحوامل اللاتي يعانين من ورم الساقين؟



زيادة الوزن، ضغط الدم، ورم الساقين

### بيانات الولادات



#### الأسئلة البحثية الرئيسية

- ما هي العوامل التي تؤثر على اختيار مكان الولادة
  - ما هي نسبة الولادات القيصرية؟
    - ما هي نسبة الولادات المبكرة؟



مكان الولادة، القائم بالولادة، نوع الولادة، مدة الحمل

#### المواليد والرضاعة



#### الأسئلة البحثية الرئيسية

- ما هي العوامل التي تؤثر على وزن المولود عند الولادة؟
  - ما هي نسبة الرضاعة الطبيعية؟ نسبة المواليد الأحياء؟.
    - توزيع نسبة المصابين بالصفراء؟
- توزيع نسبة المواليد حسب الأطوال و الأوزان بالتالي معرفة مدى حسن أو سوء تغذية الأطفال؟



#### البيانات المتوفرة

نوع المولود، وزن المولود، طول المولود، الحالة الصحية، الرضاعة الطبيعية



#### تردد الأطفال على الوحدات الصحية



#### الأسئلة البحثية الرئيسية

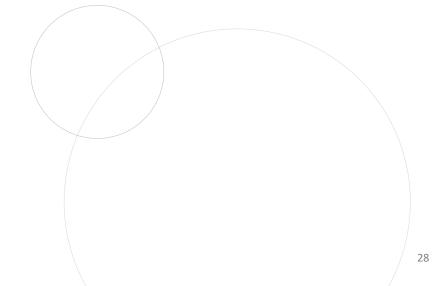
- ما هي العوامل التي تؤثر على تردد الأطفال على الوحدات الصحية
- ما هي نسبة الأطفال الذين يعانون من سوء التغذية؟
  - ما هي تغطية التطعيمات للأطفال؟



#### البيانات المتوفرة

زيارات الأطفال، وزن الطفل، طول الطفل، التطعيم، الرضاعة

# تحليل و تصوير البيانات



```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()

a_pivot = dfObj.pivot(index='Administration', columns='Year', values=values_of)

# Draw a heatmap with the numeric values in each cell
f, ax = plt.subplots(figsize=(9, 6))
sns.heatmap(a_pivot, annot=True, fmt="d", linewidths=.5, ax=ax)
```

Out[6]: <Axes: xlabel='Year', ylabel='Administration'>



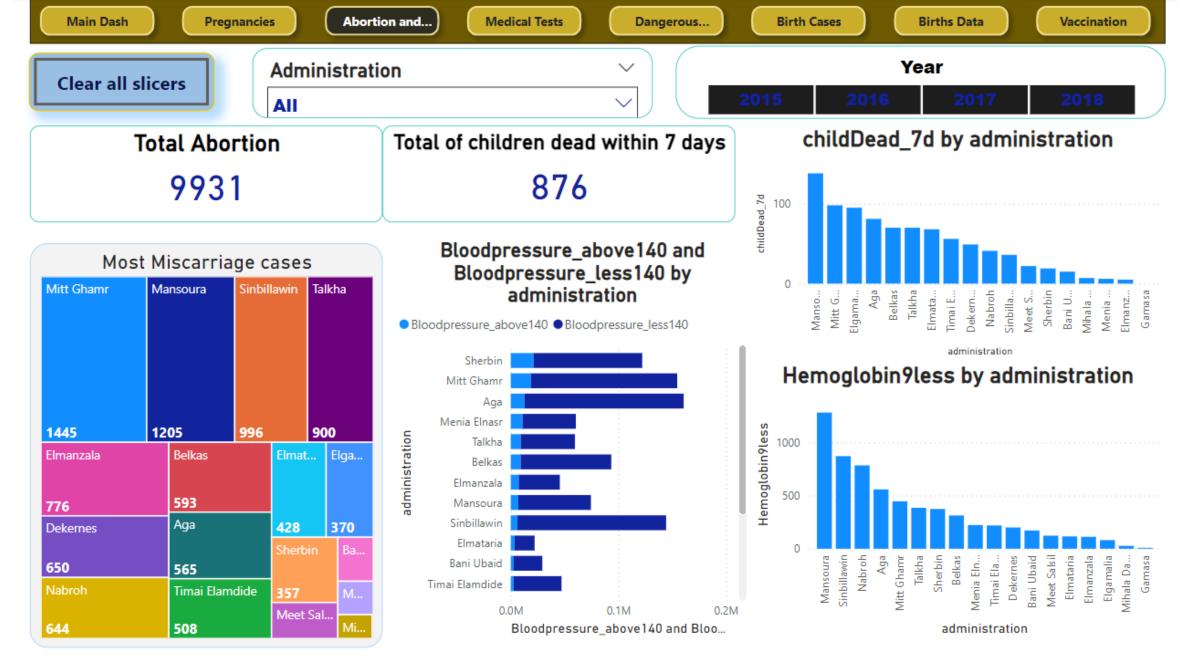


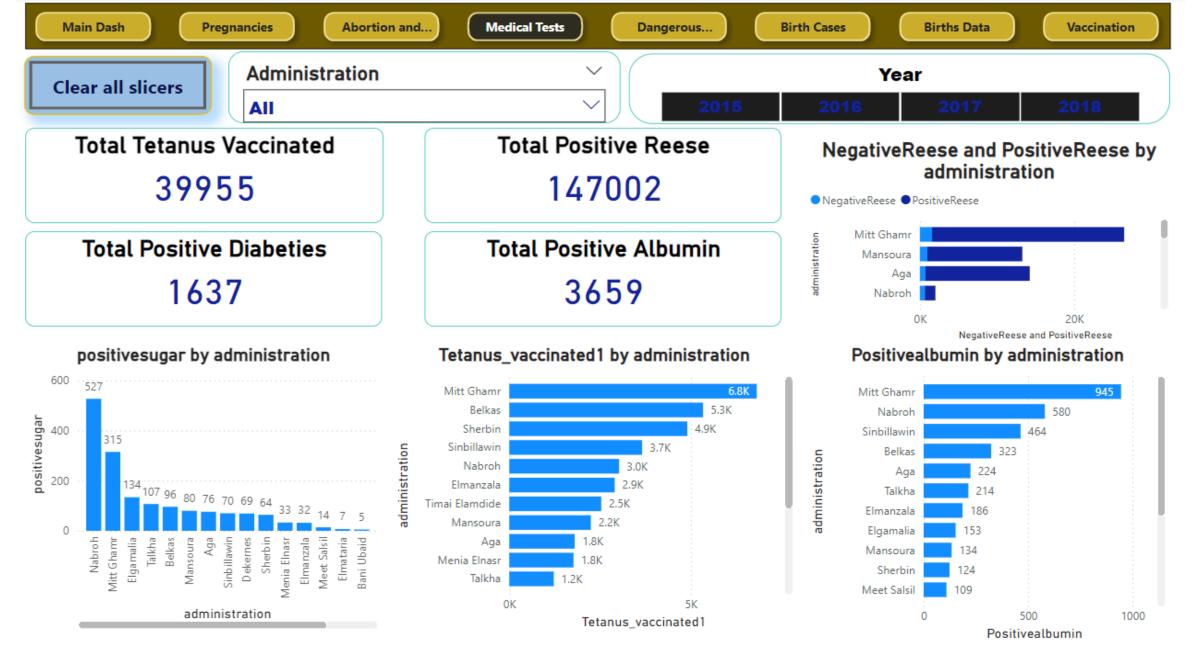
#### Healthcare Dashboard

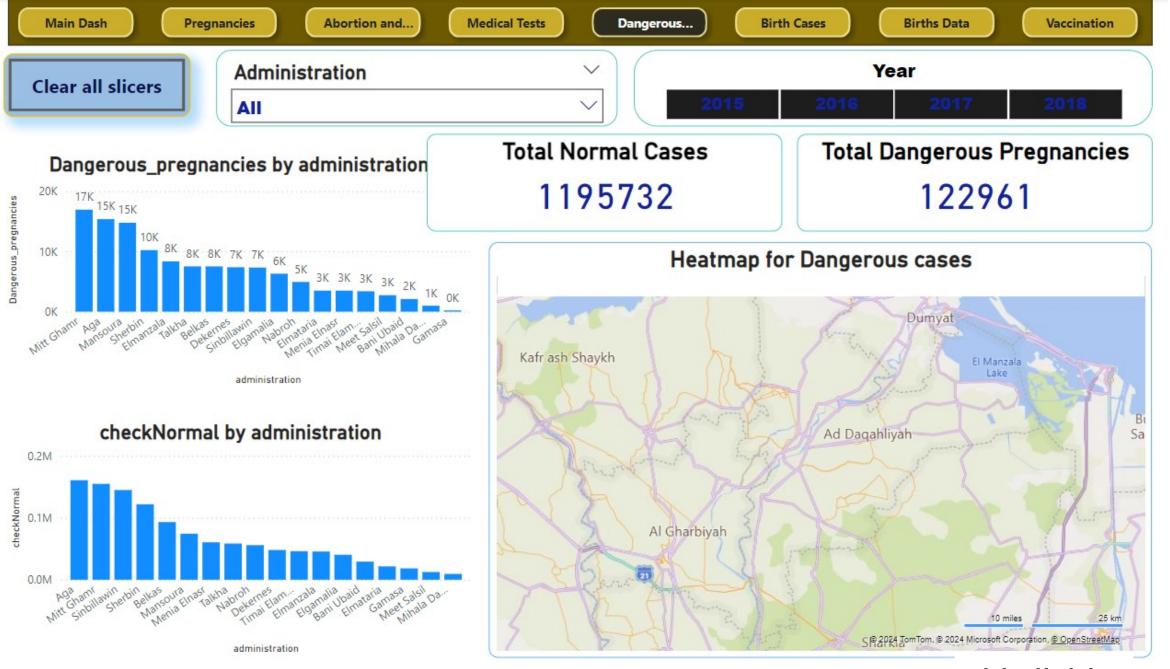


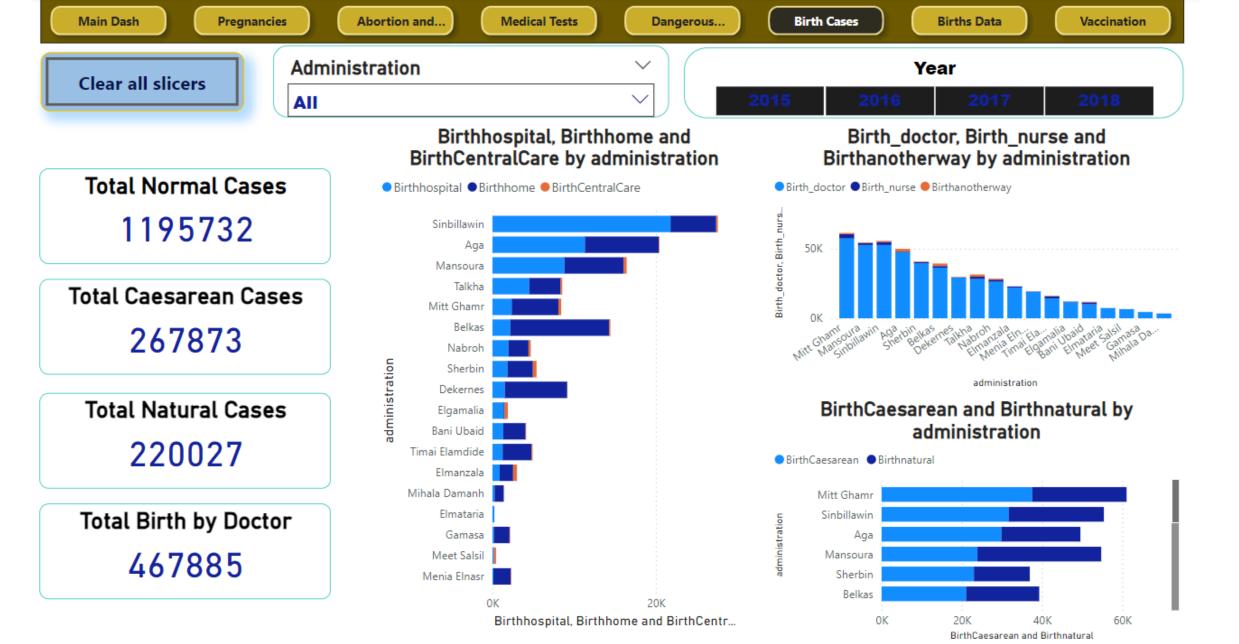




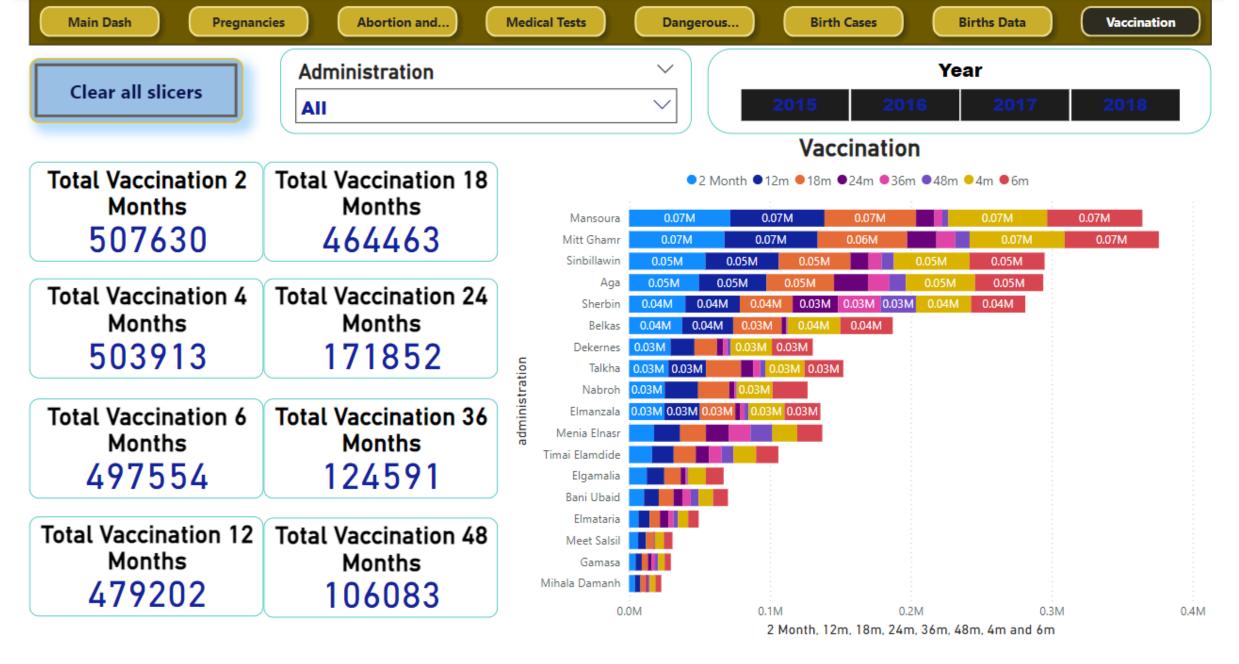
















أسعد السعدني بدير عبد الكريم

رضا إبراهيم الزكي محمد

أحمد فتحي محمود العدل

أمنية عبد العزيز إبراهيم

غاده خضر أبو بكر

اميره اشرف الجمل