

ثَمَانِيَّة

تقرير تحليل بيانات الاستماع للبودكاست

تم الإعداد بواسطة:
حصه البقمي

966507943784

Hessaalbgmi2@gmail.com

أولاً: مقدمة

تم تنفيذ هذا المشروع لتحليل سلوك المستمعين للبودكاست اعتمادًا على بيانات تتضمن:

- معلومات المستخدمين (user_id، العمر، الجنس، الدولة)

- معلومات الحلقات (episode_id، العنوان، الفئة)

- سجلات الاستماع (user_id، episode_id، مدة الاستماع)

أُجري التحليل باستخدام Python و Power BI بهدف:

- فهم الفئات الأكثر تفضيلاً.

- مقارنة سلوكيات الاستماع بين الجنسين.

- تحديد متوسط الاستماع للمستخدم الواحد.

- تقديم توصيات مخصصة وتحليل إضافي يعزز فهم سلوك الجمهور.

- ثانيًا: خطوات التحليل

- استيراد البيانات وتنظيفها

تمت قراءة الملفات باستخدام pandas، وتم حذف السجلات التي كانت مدة الاستماع فيها 0 ثانية لضمان دقة

التحليل.

- ربط الجداول

دمج البيانات باستخدام مفاتيح user_id و episode_id للحصول على جدول موحد يحتوي على:

- معلومات المستخدم

- تفاصيل الحلقة

- سجل الاستماع ومدة الاستماع

```
تحليل بيانات الاستماع للبودكاست

في هذا الدفتر، نقوم بتحميل البيانات، تنظيفها، تحليل سلوك المستخدم، وأخيرًا تقديم توصيات بناءً على الأنماط
Generate + Code + Markdown

import pandas as pd

# تحميل البيانات من المجلد
users_df = pd.read_csv("C:/Users/hessa/OneDrive/Documents/Hacker (1)_6/Desktop/8assignment/users.csv")
episodes_df = pd.read_csv("C:/Users/hessa/OneDrive/Documents/Hacker (1)_6/Desktop/8assignment/episode")
listens_df = pd.read_json("C:/Users/hessa/OneDrive/Documents/Hacker (1)_6/Desktop/8assignment/listens")

# عرض أول البيانات
users_df.head(), episodes_df.head(), listens_df.head()
```

[2] ✓ 0.2s Python

	user_id	age	gender	country
0	1	38	female	Morocco
1	2	30	female	Jordan
2	3	42	male	Saudi Arabia
3	4	57	female	Egypt
4	5	51	female	Jordan,

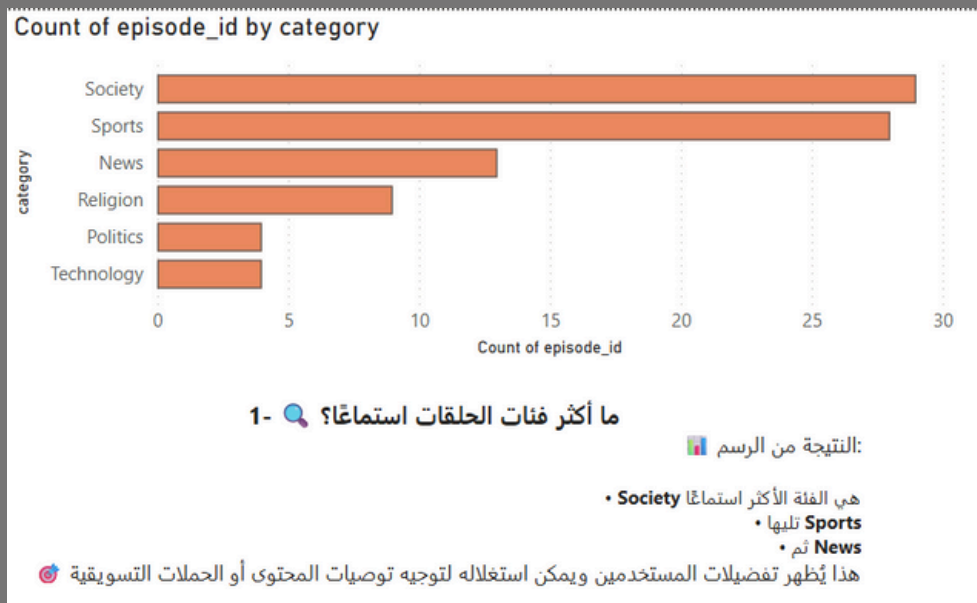
	episode_id	title	category
--	------------	-------	----------

```
listens_df = pd.read_json("C:/Users/hessa/OneDrive/Documents/Hacker (1)_6/Desktop/8assi  
# عرض أول البيانات  
users_df.head(), episodes_df.head(), listens_df.head()  
[2] ✓ 0.2s  
... ( user_id age gender country  
0 1 38 female Morocco  
1 2 30 female Jordan  
2 3 42 male Saudi Arabia  
3 4 57 female Egypt  
4 5 51 female Jordan,  
episode_id title category  
0 1 Episode 1 Religion  
1 2 Episode 2 Technology  
2 3 Episode 3 Society  
3 4 Episode 4 News  
4 5 Episode 5 Politics,  
user_id episode_id duration_seconds  
0 16 30 60  
1 6 8 120  
2 6 13 60  
3 6 27 1500  
4 17 9 1500)
```

```
# تطبيق البيانات  
listens_df_cleaned = listens_df[listens_df['duration_seconds'] > 0]  
# ربط البيانات  
merged_df = listens_df_cleaned.merge(episodes_df, on='episode_id', how='left')  
merged_df = merged_df.merge(users_df, on='user_id', how='left')  
# التأكد من القيم المفقودة  
print(merged_df.isnull().sum())  
merged_df.head()  
[3] ✓ 0.0s Python  
... user_id 0  
episode_id 0  
duration_seconds 0  
title 0  
category 0  
age 0  
gender 0  
country 0
```

```
Generate + Code + Markdown | Run All Restart Clear All Outputs | Jupyter Variables ... Python 3.11  
merged_df = merged_df.merge(users_df, on='user_id', how='left')  
# التأكد من القيم المفقودة  
print(merged_df.isnull().sum())  
merged_df.head()  
[3] ✓ 0.0s Python  
... user_id 0  
episode_id 0  
duration_seconds 0  
title 0  
category 0  
age 0  
gender 0  
country 0  
dtype: int64  
... user_id episode_id duration_seconds title category age gender country  
0 16 30 60 Episode 30 Sports 24 female Saudi Arabia  
1 6 8 120 Episode 8 Sports 59 male Morocco  
2 6 13 60 Episode 13 Society 59 male Morocco  
3 6 27 1500 Episode 27 News 59 male Morocco  
4 17 9 1500 Episode 9 Sports 21 male Saudi Arabia
```

تحليل البيانات المرئية (Power BI)



السؤال 1: ما أكثر فئات الحلقات استماعًا؟

النتيجة:

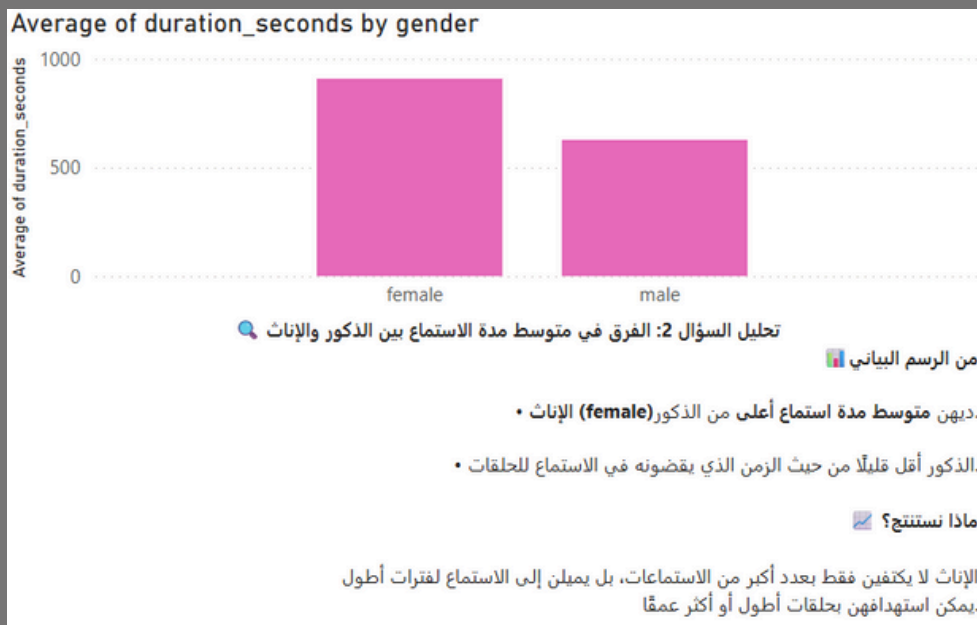
الفئة الأكثر استماعًا هي: Society

تليها:

Sports

News

هذا يعكس ميل الجمهور للمحتوى الاجتماعي والرياضي، ويمكن استغلال هذا التوجه لتوصية محتوى مخصص أو حملات دعائية للفئات الأعلى استماعًا.



السؤال 2: الفرق في متوسط مدة الاستماع بين الذكور والإناث

النتيجة:

الإناث (female) لديهن متوسط استماع أعلى

الذكور (male) يقضون زمنًا أقل في الاستماع

يُظهر هذا التباين أهمية تطوير المحتوى الذي يجذب كلا الجنسين بشكل متوازن، وقد يكون مؤشرًا على اختلاف الاهتمامات أو نوعية الحلقات المقدمة.

4.42

AverageEpisodesPerUser

متوسط عدد الحلقات لكل مستخدم: 4.42

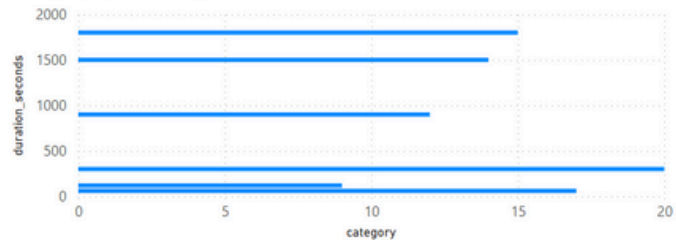
هذا يعني أن كل مستخدم يستمع تقريبًا إلى 4 إلى 5 حلقات بشكل عام.

السؤال 3: متوسط عدد الحلقات التي يستمع لها المستخدم الواحد
النتيجة:

متوسط عدد الحلقات التي يستمع لها المستخدم الواحد = 4.42

هذا يعني أن كل مستخدم يشاهد ما بين 4 إلى 5 حلقات تقريبًا، مما يدل على مستوى جيد من التفاعل، ولكنه قابل للتحسين.

category by duration_seconds



التحليل الإضافي المقترح: متوسط مدة الاستماع لكل فئة

يساعد هذا المؤشر في فهم جودة التفاعل مع أنواع المحتوى المختلفة. فمثلاً، الفئات التي يطول متوسط الاستماع لها تُعد محتوى جذاباً وفعالاً في شد انتباه الجمهور. هذا التحليل يُمكن صناع البودكاست من تعديل مدة الحلقات حسب نوع الفئة، واستهداف الفئات التي يُظهر الجمهور ولاء أعلى تجاهها من حيث الوقت.
من خلال هذا المؤشر يمكن دعم قرارات
تخطيط المحتوى
تحديد الفئات التي تستحق إنتاج حلقات أطول • تحسين التوصيات الموجهة للمستخدمين •

السؤال 4: تحليل إضافي – متوسط مدة الاستماع حسب الفئة
التحليل:

الفئات التي تمتاز بمدة استماع أطول هي مؤشر على جودة المحتوى وفاعليته.

هذه الفئات تستحق الاهتمام من حيث تخطيط المحتوى والتسويق.

من خلال هذا التحليل يمكن:

تحديد الفئات التي تستحق إنتاج حلقات أطول.

تحسين التوصيات الموجهة للمستخدمين حسب العمق الزمني.

```
def recommend_episodes(user_id: int, listens_df, episodes_df):
    user_listens = listens_df[listens_df['user_id'] == user_id]
    if user_listens.empty:
        return f"لا توجد استماعات للمستخدم {user_id}"
    top_categories = user_listens['category'].value_counts().index.tolist()
    listened_episode_ids = user_listens['episode_id'].unique()
    recommendations = episodes_df[episodes_df['category'].isin(top_categories)]
    recommendations = recommendations[~recommendations['episode_id'].isin(listened_episode_ids)]
    return recommendations.head(3)

# تجربة على مستخدم رقم 6
recommend_episodes(6, merged_df, episodes_df)
```

[4] ✓ 0.0s

Python

	episode_id	title	category
4	5	Episode 5	Politics
5	6	Episode 6	Society
8	9	Episode 9	Sports

التوصيات بناءً على تفضيلات المستخدم

تم تطوير دالة تحت اسم `recommend_episodes`، تهدف إلى اقتراح حلقات جديدة لمستخدم معيّن بناءً

على فئات الحلقات التي استمع إليها مسبقًا، وذلك عبر الخطوات التالية:

استخراج استماعات المستخدم المحدد عن طريق `user_id`.

تحليل أكثر الفئات التي استمع إليها هذا المستخدم باستخدام التكرار (`value_counts`).

استبعاد الحلقات التي استمع إليها بالفعل من قائمة الحلقات المتاحة.

اقتراح أول 3 حلقات من الفئات المفضلة لديه والتي لم يستمع لها مسبقًا.

هذه الدالة مفيدة في تعزيز تجربة المستخدم من خلال تقديم محتوى شخصي يتناسب مع اهتماماته السابقة.

✓ تطبيق الدالة على المستخدم رقم 6 أعطى توصيات لحلقات جديدة ضمن الفئات التي يفضلها ولكن لم

يستمع لها بعد.

📄 خلاصة النتائج والتوصيات النهائية

✅ النتائج المستخلصة:

الفئات الأكثر استماعًا:

جاءت فئة المجتمع (Society) كأعلى الفئات استماعًا، تليها فئة الرياضة (Sports) والأخبار (News).

يشير ذلك إلى أن الجمهور يُفضّل المحتوى الاجتماعي والرياضي بدرجة أكبر من غيره.

الفروقات بين الجنسين:

أظهرت الإناث متوسط مدة استماع أطول من الذكور.

قد يعكس هذا تفاعلًا أعمق أو اهتمامًا أكبر من الإناث بأنواع معينة من الحلقات.

متوسط عدد الحلقات لكل مستخدم:

بلغ المتوسط حوالي 4.42 حلقة لكل مستخدم.

يُشير إلى تفاعل معتدل يمكن تعزيزه بوسائل مخصصة لزيادة الاستماع.

تحليل إضافي – متوسط مدة الاستماع حسب الفئة:

بعض الفئات تحقق مدة استماع أطول، ما يدل على جودة أعلى للمحتوى أو ارتباط أقوى مع الجمهور.

🚫 التوصيات المبدئية:

🎯 زيادة إنتاج الحلقات ضمن الفئات ذات الشعبية الأعلى مثل "Society" و "Sports".

💡 تخصيص التوصيات للمستخدمين بناءً على فئاتهم المفضلة كما طُبّق في دالة recommend_episodes().

🧠 تحسين المحتوى المقدم للذكور بناءً على اهتماماتهم لتعزيز مدة الاستماع.

📈 دعم الفئات ذات متوسط استماع أطول في الحملات التسويقية وخطط إنتاج المحتوى.

🔄 تشجيع المستخدمين على استكشاف المزيد من الحلقات من خلال الاقتراحات، المكافآت، أو تقنيات الربط بين الحلقات.