**[Evaluating similarity based on correlation](https://www.linkedin.com/learning/building-a-recommendation-system-with-python-machine-learning-ai/evaluating-similarity-based-on-correlation?autoplay=true)**

تقييم التشابه على اساس العلاقه

يتم تقييم التشابه عل اساس التشابه في تقييم العملاء بحيث الاخذ بتفضل العميل للمنتجات بعين الاعتبار .

ماهو مدى الترابط بين منتجين بناءا على تقييم العملاء ؟

هنا نستخدم R-Persona لاستخراج التشابه بين منتجات قام العملاء باختيارها

بمعنى : استخراج منتج له نفس قيمة التقييم مع منتج اختاره العميل بالفعل

**{ اخترت فلم نعطيك فلم له نفس تقييمات الفلم الذي اخترته }**

**{ اخترت وجبة نعيطيك وجبة لها نفس تقييمات الوجبه اللي اخترتها }**

**{اخترت سياره نعطيك سيا ه لها نفس تقييمات السيتره اللي اخترتها }**

ماهو R-Persona :

**١ = علاقه ايجابيه قويه**

**٠ = لا يوجد علاقه**

**-١ = علاقه سلبيه قويه**

هذا النوع من التحليل سوف يطابق المنتجات بناءا على مراجعات العملاء ، من الممكن ان نجد في الداتا نس العميل قام بتقييم اكثر من منتج

**الخطوات :**

**١-تحميل المكتابات وترتيب المخرجات :**

بعد تحميل البيانات يجب ان نستخرج المخرجات التي نحتاجها منها فقط عن طريق عمل Object جديد يتضم ما نريده من كل داتا :

The Data :

fream = [ رمز العميل ، رمز المطعم ، تقييم المطعم ، تقييم الاكل ، تقييم الخدمه ]

geo = [ رمز المطعم ، نوع المطبخ ] , places = geo [[‘placeid’ , ‘name’ ]]

coisine = [ نوع المطبخ ، رمز المطعم ]

مجموع البيانات التي نريدها

{ رمز العميل ، رمز المطعم ، تقييم المطعم ، تقيم الطعام ، تقييم الخدمه ، اسم المطعم ، نوع المطبخ }

**٢- استخراج متوسط التقييمات:**

نقوم باستخراج المتوسط mean لكل تقييمات العملاء لكل مطعم ، عن طريق عمل داتا جديده تضم مجموعه من { رمز المطعم، تقييم المطعم }

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Object | new data | من داتا | method | العمودين | method |
| rating | pd.datafram | frame | groupby | (‘placeid’),[‘rating’] | mean() |

**rating = pd.DataFrame (fream.groupby(‘placeid’),[‘rating’].mean()**

**النتيجه : متوسط تققيمات العملاء لكل مطعم**

**٣- استخراج عدد التقييمات لكل مطعم :**

* كم عدد التقييمات التي حصل عليها كل مطعم ؟ وهنا نستخدم طريقة الاكثر شهرة

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| الداتا اللي عملناها | العمود الذي نريد اضافته للداتا | الداتا | من داتا | method | الحقول التي نريدها في الداتا | method |
| rating | [‘rating\_count’] | pd.dataframe | fream | groupby | (‘placeid’)[‘rating’] | count() |

**rating [‘rating\_count’] = pd.DataFrame (fream.groupby (‘placeid’) [ ‘rating’] .count()**

**النتيجه : جدول يتضمن { اسم المطعم ، التققيم ، عدد التقييمات التي حصل عليها }**

**٤- التحليل الوصفي :**

**rating.describe()**

ما يقدمه لنا :

العدد : عدد المطاعم التي تم تقييمها في الداتا اللي عملناها

ماكس : اكبر قيمة تقييم في الداتا و العدد هو عدد التقييمات

**٥- المطعم الاعلى تقييم :**

يمكننا استخراج المطعم الاعلى تقييم عن طريق جمع عدد التقييمات لكل مطعم وترتيبها تنازليا فيكون الاول هو الاعلى تقييم .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| الداتا اللي جمعنا فيها | method | العمود | تنازليا | اول 5 |
| rating . | sort\_values | (‘rating\_count’ | ascending=fals | head() |

**rating.sort\_velues ( ‘rating\_count’ , ascending = false ).head()**

**النتيجه : المطعم الاعلي تقييما مرتب بشكل تنازلي**

**٦- استخراج اسم المطعم الاعلى تقييما :**

نقوم بتطبيق فلتر يعطينا قيمة True ونمرر له العمود لسيتخرج لنا اسم المطعم

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| الداتا اللي عملنها بالاول اللي تتضمن الاسم | من هذه الداتا | من عمود | يساوي صحيحه | رمز المطعم اللي نريد اسمه |
| places | places | ‘ places id ’ | == | 1350038 |

**places [ places [‘ placeId’] == 1350038**

**النتيجه : جدول { الاندكس . رمز المطعم ، اسم المطعم }**

**٧- استخراج نوع المطبخ للمطعم الاكثر عدد تقييم**

نقوم بتطبيق فلتر لاستخراج نوع المطعم نمرر له العمود من الداتا مثل السابق

**cuisine [ cuisine [‘ placeId’] == 13582238**

**النتيجه : جدول يتضمن { الاندكس ، رمز المطعم ، نوع المطبخ }**

**٨- تجهيز البيانات للتحليل :**

- عمل مصفوفة تقييمات العملاء :

مصفوفة تققيمات العملاء هي جدول يوضح تقييمات العملاء بالكامل في كل الداتا وتشير الى تقييم كل عميل

امام المطعم الي قام بتقييمه بقيمة التقييم ويستخدم Poivot Tabel لعرض المصفوفه .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Object | method | من داتا | العمود | اسم العمود الجديد |
| places\_crostab | pd.poviot\_table | (data=fream | values=’rating’ | columns=places |

**places\_crosstab = pd.povit\_table ( data = fream, values = ‘rating’ , columns = ‘ places’ )**

**النتيجه : جدول يتضمن الكثير من NAN وهذا بسبب ان ليس الجميع قاموا بتقييم جميع المطاعم ، ويتضمن الجدول { التقييم ، رمز المطعم ، رمز العميل } القيمه موجده فقط عند المطعم الذي قام العميل بتقييمه .**

**٩- ايجاد الاماكن ذات العلاقه :**

١- يجب ان نقوم بحذف البيانات التي تتضم NAN وذلك عن طريق انشاء داتا جديده تتضمن مانرديه فقط وهو مطعمنا .

|  |  |  |
| --- | --- | --- |
| Object | الداتا اللي عملناها سابقا ٠ من هذه الدتا | رمز المطعم الاكثر شهره ، مطعمنا |
| Tortas\_rating | places\_crosstab | 1385002 |

٢- نقوم بعمل فلتر لكي يضهر لنا التقييمات الخاصه بمطعمنا فقط واستخراج التقييمات الي اعلى من / تساوي الصفر

**Tortas\_rating [Tortas\_rating >= 0 ]**

**النتيجه : يضهر لنا جميع التقييمات التي اكبر من او تساوي صفر ( جميع التقييمات بالتفصيل )**

**١٠- لايجاد علاقه التشابه :**

سوف نستخرج الان العلاقه بين مطعمنا والمطاعم الباقيه لنعرف الاكثر تشابهها عن طريق استخراج R-Persona corrlation

|  |  |  |  |
| --- | --- | --- | --- |
| Object | الداتا اللي قمنا بعملها سابقا | method | الداتا التي نريد المقارنه بها |
| similer\_to\_tortas | places\_crosstab | corrwith | (tortas\_rating) |

similer\_to\_tortas = places\_crosstab , corrwith ( tortas\_rating)

النتيجه : سوف يستخرج لنا مصفوفه نقوم بتحويلها الى داتا فريم

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Object | new Data | الداتا اللي عملناها | اضافه العمود | اسم العمود الجديد للاضافه |
| corr\_tortas | pd.dataframe | (similer\_to\_torta | colmuns | [‘pearsonsR’]) |

**corr\_tortas = pd.DataFrame ( similer\_to\_tortas , columns = [‘ pearsonR’]**

**النتيجه : جدول يتضمن {رمز المطعم ، Pearsons R } ، القيم في همود Persona R هي قيمة التقييم R-Persona , العلاقه بين كل المطاعم مع مطعمنا .**

**ملاحظة مهمه جدا :**

**يمكن ان نحد مطعم له نفس العلاقه مع مطعمنا ولكنه يتضمن عدد قليل من التقييمات . لذلك وصل الى هذه العلاقه وهذا غير صحيح .**

**وهنا يجب ان ناخذ بالاعتبار مدى شهرة المطعم { ليس اي مطعم صغير يشابه مطعمنا }**

**١١- لاستخراج المطاعم ذات التقييمات المناسبه :**

نقوم بدمج الداتا مع داتا التقييم لكل المطاعم وهي rating

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Object | الداتا السابقه | method | من الداتا | دمج في العمود |
| tortas\_corr\_summry | corr\_tortas. | join | (rating | [‘rating\_count’]) |

**tortas\_corr\_summary = corr\_tortas.join (rating [‘ rating\_count’] )**

الان سنقوم بانشاء فلتر يتضمن من هذه الداتا المطاعم الحاصله على اكثر من 10 تقييمات وترتيبه اتنازليا

**tortas\_corr\_summury ( tortas\_corr\_summery [‘ rating\_count’] >=10 ).sort\_values (‘ personaR’ , ascednce = False ).head()**

**النتيجه : جدول يتضمن { مز المطعم ، تقييمم R ,عدد التقييمات اكثر من / تساوي 10 }**

**الان بعد هذه النتيجه نلاحظ ان هناك مطاعم لديها persona R = 1**

**وهي ليست ذات معنى لانها تعنب ان هناك العميل مكرر ، اي انه كان هناك عميل واحد قام بمراجعة كلا المطعمين ( مطعمنا ومطعم اخر ) واعطى نفس النتيجه لكلا المطعمين**

**١٢- حذف المخرجات القليله :**

سنقوم بحذف الاماكن التي تقييماتها قليله عن طريق عمل داتا فريم جديده تتضمن الباقي فقط (7) من ضمنهم مطعمنا .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Object | new Data Fram | رمز ال ٧ مطاعم التي نريدهم | اندكس بعدد الصفوف | اسم العمود في الداتا الجديده |
| places\_corr\_tortas | pd.datafream | ([ id ,id ,id, ..] | index = np.arrange(7) | colmuns [‘places’] |

**places\_corr\_tortas = pd.DataFram ([id,id,id,..] , index = np.arrang(7) , colmuns [ ‘placeId’])**

**١٣- دمج الداتا المستخرجه مع الدتا التي تتضمن نوع المطبخ :**

نقوم بدمج الداتا لكي نعرف نوع المطبخ لكل مطعم وايجاد المشابه لمطبخ مطعمنا

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Object | method | الداتا عملناها سابقا | داتا التي تتضمن نوع المطبخ | الدمج في عمود |
| summury | pd.merge | (places\_corr\_tortas | couisine | on=‘placeId’) |

**summury = pd.merge (places\_corr\_torats , couisine , on = ‘placeId’ )**

**النتيجه : جدول يتضمن { اندكس ، رمز المطعم ، نوع المطبخ }**

**عندما نطبع هذا الكود نحصل على ٥ نتائج فقط مع اننا قمنا بادخال ٧ مدخلات والسبب هو :**

**انه لم يتم ادراج جميع المطاعم في الدتا coisine**

**مع ذلك :**

**ان ما نراه هنا انه من بين افضل 6 مطاعم مرتبطه بمطعمنا يوجد على الاقل واحد لديه نفس المطبخ**

* **الان سنقوم بتطبيق فلتر للحصول على اسم هذا المطعم المرتبط مع مطعمنا**

**places ( places [ ‘ placeId’] = 12500023 )**

**النتيجه : اسم المطعم الذي يرتبط وله نفس العلاقه مع مطعمنا**

**١٤- التقييم :**

لتقييم مدى ملائمة مقياس التشابه حقا ، يجب الاخذ بموجموعة الاحتمالاتباكملها مثل عدد المائكولات في قائمه الطعام

لذلك سوف نقوم بالتحليل الوصفي :

**couisine [‘ Rcousine ‘ ].describe()**

**النتيجه :**

**العدد : عدد المطابخ بالكامل**

**يونيك : عدد المطابخ الاصلي**

**توب : افضل المطابخ**

**فريك : التردد**

**النتيجه التحليله :**

**حصلنا على 6 من افضل المطاعم التي كانت متشابهه مع مطعمنا على اساس الارتباط والشهره .**

**من بين هذه المطاعم يوجد مطعم واحد يقدم نفس مطبخنا .**

**وهنا يبدو ان نضامنا يعمل بشكل رائع في بناء التوصيه**

**والان نستطيع ان نقدم التوصيه للعملاء مطعمنا بزيارة المطعم الجيديد الذي تم استخراجه في الجدول .**