SQL

1. Į MYSQL Workbench iš Sales.csv įsikėliau duomenis, formatavau ir k.t.

2. SQL kodo pagalba susikūriau dimensines lenteles, kurios saugo informaciją apie customers, orders, products ir faktinę pardavimų lentelę.

```sql
CREATE TABLE DIM_Products (
    Product_ID VARCHAR(20) PRIMARY KEY,
    Category VARCHAR(50),
    Sub_Category VARCHAR(50),
    Product_Name VARCHAR(255)
);

INSERT INTO DIM_Products (Product_ID, Category, Sub_Category, Product_Name)
SELECT DISTINCT
    Product_ID,
    Category,
    Sub_Category,
    Product_Name
FROM Sales;
```

3. Eksportavau lenteles į aplanką „final dataacademy“ sql text file ir csv formatu, nes dar galvoju, kuriuo būdu bus patogiau naudoti POWER BI 😊)

4. **DIM_Customers** lentelėje saugoma visa informacija apie klientus.

DIM_Products lentelėje saugoma informacija apie produktus. **DIM_Orders** lentelėje saugoma informacija apie užsakymus. **FACT_Sales** lentelė susieja visus dimensinius duomenis ir saugo faktines pardavimų reikšmes.

POWER BI

1. Atsidarome BI desktop, ir pasirenkame duomenų įkėlimą per MySql duomenų bazę

2. Įsikėliau duomenis :

3. Importavus pereinu į MODEL , patikrinu ryšius tarp lentelių , tikėjausi kad bus sukurti automatiškai, beeet dėja ☹️ susimodeliavau ir susijungiau visas lenteles ryšiais ! :D

DIM_Customers Customer_ID -> FACT_Sales Customer_ID DIM_Products Product_ID -> FACT_Sales Product_ID DIM_Orders Order_ID -> FACT_Sales Order_ID.

4. Peržiūriu ir transformuoju duomenis, skaitinėmis, textinemis ir decimal reiksmemis kur reikia

5. Prieš pradedant duomenų vizualizavimą susikūriau failą word “User stories” , kuris padės man susigaudyti, kokios srities spec kokių +- užduočių tikisi iš analitiko ir ką svarbu atvaizduoti.

Kiti nori žinoti

What is user story in data analytics?

Summary: A user story is **an informal, general explanation of a software feature written from the perspective of the end user**. Its purpose is to articulate how a software feature will provide value to the customer.

ID	User Story	Acceptance Criteria	Notes
1	Kaip duomenų analitikas, noriu sukurti pardavimų prognozavimo modelį, kad galėčiau tiksliai numatyti būsimus pardavimus, remdamasis istoriniais duomenimis.	Pardavimų prognozavimo modelis turi būti sukurtas naudojant istorinius pardavimų duomenis. Modelis turi pateikti prognozes su bent 80% tikslumu.	Naudoti mašininio mokymosi algoritmus, pvz., ARIMA, LSTM, ar regresiją.
2	Kaip pardavimų vadybininkas, noriu sužinoti, kurios produktų kategorijos generuoja didžiausią pelną ir pardavimus, kad galėčiau optimizuoti prekių asortimentą.	Ataskaita, rodanti skirtingų produktų kategorijų pardavimų apimtis ir pelną. Ataskaita turi būti lengvai skaitoma ir interaktyvi.	Naudoti vizualizacijos įrankius, tokius kaip Tableau arba Power BI.
3	Kaip rinkodaros specialistas, noriu suprasti sezoniškumo tendencijas, kad galėčiau planuoti rinkodaros kampanijas.	Sezoniškumo analizė turi būti pateikta grafiškai, rodant pardavimų pokyčius skirtingais metų laikais.	Įtraukti mėnesinius pardavimų duomenis, sezoninius pokyčius.
4	Kaip verslo analitikas, noriu segmentuoti klientus pagal jų pirkimo elgseną, kad galėčiau geriau pritaikyti pasiūlymus ir akcijas.	Klientų segmentavimo analizė, remiantis jų pirkimo istorija ir elgsena. Turi būti sukurti bent 3 skirtingi segmentai.	Naudoti klasterizacijos metodus, pvz., K-means.
5	Kaip finansų analitikas, noriu įvertinti, kaip kainų pokyčiai įtakoja produktų pardavimus, kad galėčiau padėti nustatyti optimalias kainas.	Kainų elastingumo analizė, rodanti, kaip skirtingi kainų pokyčiai įtakoja pardavimus ir pelną.	Naudoti regresijos analizę.

6. ID1 – sales forecast pasilieku PYTHON analizei, ID2 – pardavimų vadybinkui pamėginsiu sukurti dashboard’e vieną puslapį, ID3 – Rinkodarininkui – pardavimai pokyciai ir pan. ID4 – klientu segmentavimas, bandysim igyvendinti 😊 nes dar nezinau kaip tai padaryti :D :D ID5 – manau naudosim PYTHON. Bet matysime eigoje 😊

7. **PRADEDAM !** 😊

8. PLANAS –

titulinis padaryti nuorodas į puslapius per rodykles? (žalias rodykles) , + įkelti piešinėlį atitinkantį sales report ‘įmonės ‘ logo ar kažką pan.
PRODUCT CATEGORY PERFOMANCE – kolkas tuščias, tik idėja pavadinimui pagal useers stories, bet idėja pardavimai pagal produktu kategorijas, pardavimu kategoriju dalys? Gal pie charte, kad filtruojant matytusi skirtumas, pardavimai pagal subkategorijas ir svarbiausia filtrai. Metai menesiai , gal sugalvosiu kokius kpiaisus
SESONALUTY TRENDS - Pardavimų apimtys per metus, Pardavimų aktyvumas per mėnesius, DAR GALVOSIM 😊 judam nuo pradziu, kuriam title

9. Yep It’s ME :D pradėjau daryti nuo paskutinio puslapio :D :D klientų segmentacijos, radau internete info, Klientų segmentavimas gali būti atliekamas naudojant įvairius metodus, vienas iš populiariausių yra RFM analizė (Recency, Frequency, Monetary). Bandysiu susikurt kažką, naujus stulpelius, ar measursus, kad kažkaip juos išfiltruočiau :D

10. Prikūriau stulpelių 😊 susegmentavau klientus !!!! reikia sugalvoti, kaip visa tai vizualizuoti

11. Ir pagaliau baigiau pirmą (paskutinį) puslapį :D Pritaikiau vizualą, kurį manau naudosiu, nes patinka 😊

12. Noriu sales sukurti comparison, kad rodytų rodyklėlę interaktyviai nuspaudus, kolkas nežinau kaip tai padaryti, bet jau džiaugiuos kad verslo analitikui jau pateikiau segmentuotus klientus :D woop woop :D

13. Prasideda bėdos :D 😟 (((((((nenuskaito dim_products lenteles, nors sujungta ryšiais, stulpeliai sutampa ir pan :D rodo tuščias reikšmes, nors tablewiev viskas ok 😊) wish me luck :D

14. Ištrinu lentelę, ir keliu duomenis iš naujo 😊

15. VEIKIAAAAAAAAAAAA!!

16. Dar vienas puslapis baigtas !!! 😊))

17. Day 3 ! 😊 pradedam trečią ataskaitos ruošimo dieną, su nauju puslapiu seasonaly trends, bandysime, kaip pavyks (turi pavykt) ir kiek užtruks.

18. Viskas, paskutinį puslapį pasidariau per 20 min :D nesuvokiu, kaip taip greitai, beeet viskas sklandžiai ir aiškiai :D

19. Paskui dar norėčiau su tooltip pakoreguoti duomenis, + titulinį pagražinti 😊

PYTHON

1. Pradedam phyton analizę, viską bandysiu atlikti per Visual Studio code programą Jupyter notebooke . įsikėliau aplanką, susiprašiau team, probleminį klausimą. +- potemes ką galiu veikt, beeet manau, kai pradėsiu analizę, dar viskas keisis 😊)))

2. Bandom 😊) kolkas bandžiau su 100 estimatorių, beeet 25 min neužkrovė modelio, tikėtina dėl duomenų kiekio, modelio hiperparametro, t.y. šakų sk. Arba labiausiai tikėtinas mano kompiuteris :DD

3. Todėl bandžiau sumažinti n_estimators kiekį 20, taip modelį sugeneravo per 8m su trupučiu.

Mean absolute error

Mean absolute error, abbreviated as MAE, is a metric used to measure the average absolute difference between the predicted and actual values in a regression problem i.e. modeling various relations of variables.

4. Woop wooop , trukt už vadžiu vėl iš pradžiu :D :D MAE – kiek prognozuojamas modelis skiriasi nuo realių reikšmių, tai realiai kuo mažesnis skaičius tuo geriau modelis atlieka prognozę.. šiuo atveju reikalai nekokie, ir tokiam modeliui nėra šansų 😊 bandom kažką kitką..

```python
# Įvertiname modelį
from sklearn.metrics import mean_absolute_error

mae = mean_absolute_error(y_test, predictions)
print(f'Mean Absolute Error: {mae}')
```

[34] ✓ 0.0s Python

```
Mean Absolute Error: 141.48760782771535
```

5. Išbandžiau dėja, MAE = 248

```python
model = GradientBoostingRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

6. Bandau tiesinę regresiją MAE = 189, vis dėl to geriausiai randomforest beet ištryniau jos kodą :D :D :D :D

7. Googlinu, kiek yra ML regresijų :D :D ir išbandžiau Lasso, tai MAE 184, kaži ar neliksime ties RandomForest :D :D SWR = 229 catboost=209

8. Nusprendžiau likti prie randomforest, pabandysiu iš naujo sugaudyti kodą. Ir treniruoti modelį. Bandau treniruoti, užtruko 39 minutes 😊 gavau fitfailedwarning – kaikurios treniruotės nepavyko ☹️ kodėl? Gali būti per daug platus hiperparametrų rinkinys, max_features parametron klaida auto reikšmė.

Modelio minusas ilgas treniravimo laikas, bandysiu mažint hiperparametro reiksmes, plius keisti, kad pasalintu auto reiksmes

9. Modelis pasiekė MAE 152.11, kuris yra geras rezultatas ir rodo, kad modelis gana tiksliai prognozuoja pardavimus. 😊)) nor sir nelabai džiugina, bet geriausia ką šiuo atveju galiu padaryti. (gal peržiūrint duomenis ir leikanas šalinant būtų galima kažką koreguot)

10. Tikrinant kaip klaidos pasiskirsto pagal laiką, neatrodoo labai blogai 😊 visada norisi geriau, beeet :D

11. Pavargus,ir nebežinau ką daryt :D :D bet gal u-žteks, padarysiu išvadas ir pateiksiu python analizę.

Untitled-1.ipynb
Code
Markdown
Run All
Restart
Clear All Outputs
Variables
Outline
Python 3.12.3
plt.show()
[91]
0.9s
Python
Prognozuoti pardavimai pagal produktų subkategorijas
Prognozuoti pardavimai
800000
700000
600000
500000
400000
300000
200000
100000
0
Accessories
Appliances
Art
Binders
Bookcases
Chairs
Copiers
Envelopes
Fasteners
Furnishings
Labels
Machines
Paper
Phones
Storage
Supplies
Tables
Produkto subkategorija