Fiktívny podnik - Bageta, s.r.o.

Projekt z predmetu Business Intelligence

Alena Stracenská & Juraj Šiandor

november 2022

Opis firmy Bageta, s.r.o.

Firma "Bageta s.r.o." je fiktívna spoločnosť vymyslená autormi tejto semestrálnej práce. Autori tejto práce sa dištancujú od akejkoľvek obchodnej spoločnosti, ktorá bola v minulosti, je v súčasnosti alebo bude v budúcnosti zapísaná pod týmto alebo podobným obchodným názvom v obchodnom registri SR. Všetky dáta, vrátane adries prevádzok sú fiktívne a nereprezentujú žiadny obchodný alebo právny vzťah ku spoločnostiam alebo právnickým osobám podnikajúcich na týchto adresách.

Firma Bageta, s.r.o. začala písať svoju históriu v roku 2010. Na začiatku bola vedená ako podnik s dvoma zamestnancami, no neskôr sa rozrástla až do súčasnej podoby. Zameriava sa na predaj tradičných pekárenských výrobkov chleba, rožkov a pleteniek, ale nedávno pribudli do jej portfólia aj ďalšie nové bezlepkové produkty a sladké pečivo. Firma má viacero predajní po západnom Slovensku.

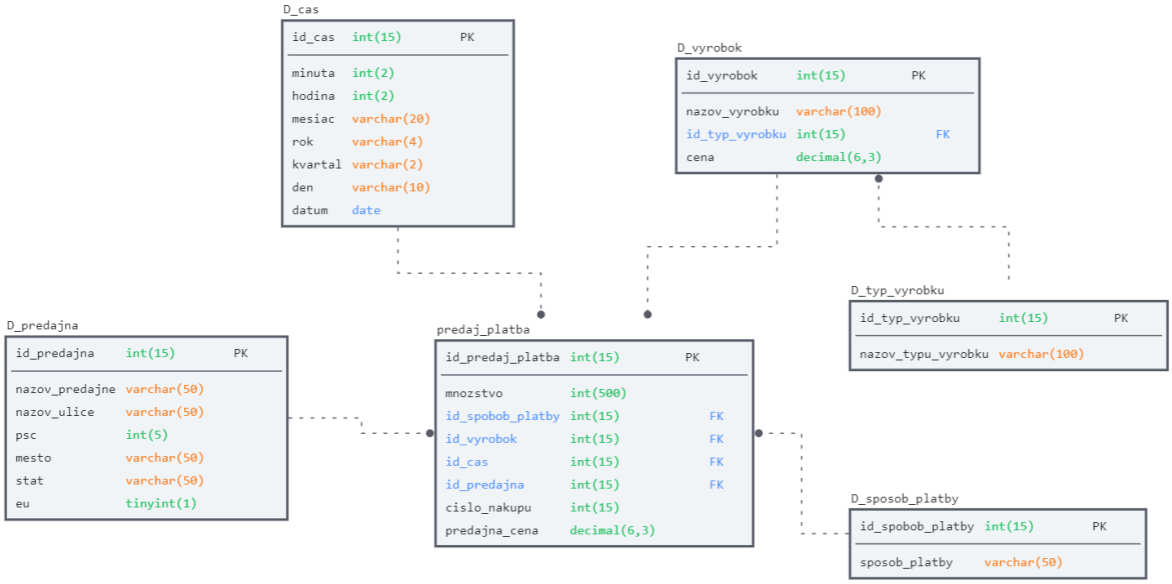
Cieľom tohto projektu je vytvoriť dátový sklad pre našu firmu. Neustále pribúdanie nových typov výrobkov a zákazníkov nás motivuje mať toto kvantum dát uskladnené a analyzované. Chceme mať prehľad o tom, v akej predajni sa toho najviac predá a v akom čase sa predáva tovar najviac, prípadne aký tovar sa predáva najviac počas jednotlivých mesiacov.

Hlavným dôvodom, prečo chce naša firma zaviesť dátový sklad je pre uchovávanie veľkého množstva súčasných aj budúcich dát a takisto pre efektívne uskladnenie informácií na jednom mieste, práve to dátové sklady umožňujú.

Na aké otázky by sme chceli po zavedení dátového skladu vedieť odpoveď?

- V akej predajni a v akom čase sa predáva tovar najviac ?
- Aké výrobky sa predávajú najviac v lete, a aké v zime?
- Aká je priemerná cena nákupu na všetkých predajniach v jednotlivých krajoch?
- Ktoré výrobky vykazujú najnižší a najvyšší zisk za jednotlivé mesiace podľa predajní?
- Akou formou najčastejšie platia zákaznici v jednotlivých prevádzkach kartou alebo hotovosťou (zvýhodnenie zmluvných podmienok s bankou o prenajatí terminálu) ?
- Aké máme tržby za jednotlivé predajne v priebehu rokov ?

Veríme, že zavedením dátového skladu budeme vedieť zodpovedať na vyššie uvedené otázky v odrážkach a zároveň dúfame, že výsledky, ktoré dostaneme budú pre nás priaznivé z dôvodu, že v ďalších rokoch by sme radi otvorili predajne aj v susedných krajinách. V logickom modeli sa budeme snažiť namodelovať, ako by mohol daný dátový sklad vyzerať. Budeme ho robiť v tvare "snowflake", čiže snehovej vločky.



Obrázok č. 1: Logický model

Kvôli MySQL sme museli zmeniť atribút **mnozstvo** vo faktovej tabuľke **predaj_platba** z 500 na 225.

ETL proces

V ETL procese sme z dimenzie **D_vyrobok** vyhodili atribút **typ_vyrobku** a pridali atribút **cena**. Následne sme si manuálne naplnili 3 dimenzie **D_typ_vyrobku**, **D_vyrobok** a **D_predajna**, kvôli tomu, že nám bolo jednoduchšie doplniť 30 hodnôt manuálne ako špecificky zisťovať, či na internete existuje generátor dát na generovanie predajní po západnom Slovensku, prípadne či vieme niekde vygenerovať typy pečiva.

Zvyšné 2 dimenzie **D_cas** a **D_sposob_platby** a faktovú tabuľku **predaj_platba** sme generovali prostredníctvom Pythonu.

V prvom kroku sme spravili spojenie na databázu a overili jeho funkčnosť.

```
#pripojenie sa na databazu a overenie spojenia
import MySQLdb
import pandas as pd
import numpy as np
import random
from datetime import datetime
from faker import Faker
mydb = MySQLdb.connect("localhost", "root","", "pekaren")
mycursor = mydb.cursor()
mycursor.execute("Show databases")
random.seed(datetime.now())
for db in mycursor:
    print(db)
```

V druhom kroku sme načítali .csv súbor, ktorý po načítaní vyzeral takto:

```
#nacitanie udajov z csv
#manualne sme si k zakladni vyrobkov pridali aj priemerny pocet, standardnu odchylku a
    maximum,
#kolko zakaznik v realnom svete takeho tovaru aj kupi
##tzn. ze na 1 nakup nekupi 8 chlebov a 1 rozok ale 1 chlieb a 8 rozkov +- nejaka rozumna
    odchylka.
#na generovanie poctu kusov pouzijeme Gaussa
df = pd.read_csv("d_vyrobok.csv", sep = ';')
```

```
df.head(20)
```

	id_vyrobok	nazov_vyrobku	id_typ_vyrobku	cena	priemer	std	max
0	1	chlieb pšenično-ražný konzumný	1	1.49	1	1	4
1	2	slatinský chlieb zemiakový	1	2.49	1	1	4
2	3	chlieb ražný	1	1.59	1	1	4
3	4	chlieb tradičný kváskový	1	1.59	1	1	4
4	5	chlieb ražno-špaldový	1	2.85	1	1	4
5	6	rustikálny chlieb z kamennej pece	1	2.15	1	1	4
6	7	pastiersky chlieb bez E	1	2.05	1	1	4
7	8	chlieb sedliacky	1	2.79	1	1	4
8	9	baker street kváskový semienkový chlieb	1	3.59	1	1	4
9	10	rožok štandard	1	0.13	8	3	24
10	11	kajzerka natural	1	0.12	8	3	20
11	12	kajzerka cereálna	1	0.17	8	3	20
12	13	bageta francúzska	1	0.28	3	2	8
13	14	kornbageta	1	0.39	3	2	8
14	15	pletenka	1	0.25	4	2	10
15	16	rožok grahamový	1	0.30	8	3	24
16	17	mini pizza šunková s kukuricou	1	0.69	2	1	8
17	18	závitok s pudingom a hrozienkami	2	0.55	2	1	8
18	19	rožtek s pudingovou náplňou	2	1.49	2	1	8
19	20	makový mini závin	2	0.65	3	2	8

Obrázok č. 2: Obsah .csv súboru

V treťom kroku už vygenerujeme náhodnú predajňu, počet kusov pečiva, čas a všetko ostatné potrebné pre doplnenie zvyšných tabuliek.

```
#Vygeneruje pocet kusov daneho vyrobku na zaklade priemeru a standardnej odchylky zadanej manualne ku kazdemu vyrobku
#Max sluzi hlavne na to aby sa vo vygenerovanych hodnotach z casu na cas nenasiel outlier co kupi napr 456789876 rozkov...
#Metoda ale nerata ze v ponuke mame 5 druhov chlebov a zakaznik pri volbe poloziek "kupi" 5 roznych chlebov po 1 kus
def generate_pocet(mean,sigma,_max):
    ret=round(random.normalvariate(mean, sigma))
    while ret<=0:
        ret=round(random.normalvariate(mean, sigma))
    return ret if ret<=_max else _max

#https://stackoverflow.com/questions/553303/generate-a-random-date-between-two-other-dates (ciastocne)
#vygenerujeme semi-nahodne datumy v nejakom rozmedzi (end_date +n rokov znamena ze hodnota roku bude +n, nie ze +n od daneho datumu)
```

```
#cas drzime v pracovnej dobe (6-18 hod), "otvorene" je ale 365 dni v roku,nezohladnujeme sviatky
def generate_datetimes(m):
    dt=[]
    for i in range(m):
        fake = Faker()
        import datetime
        start_date = datetime.date(year=2016, month=1, day=1)
        datetime=fake.date_time_between(start_date=start_date, end_date='+0y')
        if datetime.hour < 6 or datetime.hour >= 18:
            datetime=datetime.replace(hour=random.randint(6,17))#random.randint(6,17)
        kvartal=datetime.month//4+1
        dt.append([datetime.minute,datetime.hour,datetime.month,datetime.year,kvartal,datetime.day,datetime])
    dt=np.array(dt)
    dt=dt[np.argsort(dt[:,6])]
    return dt

m=75000
dt=generate_datetimes(m)
id=1
for predaj in range(1, m):
    #nahodna predajna
    predajna=random.randint(1,13)
    #pocet poloziek (generuje od nuly preto +1, exponencialne rozdelenie,
    #cim mensia lambda tym vacsie cisla, najvacsiu pravdepodobnost maju male cisla)
    pocPol=round(random.expovariate(0.8))+1
    #platba kartou -> 0, platba v hotovosti -> 1
    spPlatRnd=random.randint(1,100)
    spPlat=1 if spPlatRnd < 66 else 0
    #print("predaj:",predaj,",predajna-id:",predajna,",poc poloziek:",pocPol,":")
    #vklad datumov
    sql = "INSERT INTO d_cas (id_cas, minuta, hodina, mesiac, rok, kvartal, den, datum) VALUES (%s, %s, %s, %s, %s, %s, %s, %s)"
    val = (predaj,dt[predaj-1,0],dt[predaj-1,1],dt[predaj-1,2],dt[predaj-1,3],dt[predaj-1,4],dt[predaj-1,5],dt[predaj-1,6])
    mycursor.execute(sql, val)
    #vklad platieb
    sql = "INSERT INTO d_sposob_platby (id_spobob_platby, sposob_platby) VALUES (%s, %s)"
    val = (predaj, spPlat)
    mycursor.execute(sql, val)

    mydb.commit()
    # ID casu, id predaj_platba a id platby su tie iste hodnoty, incrementy
    for polozka in range(pocPol):
        #vyberieme nahodny vyrobok
        vyrobok=random.randint(0,34)
        pocet=generate_pocet(df['priemer'][vyrobok],df['std'][vyrobok],df['max'][vyrobok])
        cena=pocet*df['cena'][vyrobok]
        #print(vyrobok,",",df['nazov_vyrobku'][vyrobok],",",pocet)
        sql = "INSERT INTO predaj_platba(id_predaj_platba, mnozstvo, id_spobob_platby, id_vyrobok, id_cas, id_predajna, cislo_nakupu, predajna_cena) VALUES (%s, %s, %s, %s, %s, %s, %s, %s)"
        val=(id,pocet,predaj,vyrobok,predaj,predajna,predaj,cena)
        mycursor.execute(sql, val)
        mydb.commit()
        id+=1
```

```
#vymazanie udajov z DB
sql = "DELETE FROM predaj_platba"
mycursor.execute(sql)
sql = "DELETE FROM d_sposob_platby"
mycursor.execute(sql)
sql = "DELETE FROM d_cas"
mycursor.execute(sql)
mydb.commit()
```

SQL scripty

Do databázy pekárne sa dostaneme cez jazyk R, z ktorého sme schopní exekuovať skripty a vidieť výsledky dopytov prehľadne zobrazené.

```
#pripojenie kniznice do projektu
library(RMySQL)
```

```
## Loading required package: DBI
```

```
mysqlconnection = dbConnect(RMySQL::MySQL(),
                            dbname='pekaren',
                            host='localhost',
                            port=3306,
                            user='root',
                            password='')
#zobrazenie vsetkych tabuliek v db pekaren
dbListTables(mysqlconnection)
```

```
## [1] "d_cas"           "d_predajna"      "d_sposob_platby" "d_typ_vyrobku"
## [5] "d_vyrobok"       "predaj_platba"
```

Otázka č. 6 - Aké máme tržby za jednotlivé predajne v priebehu rokov ?

Interpretácia: Za jednotlivé predajne v jednotlivých mestách a adresách máme tržby zoradené od najvyššej po najnižšiu v nasledujúcej tabuľke. V podstate vidíme, že najvyššie tržby máme vo Svätom Jure a keďže je tam iba jedna predajňa, mohli by sme uvažovať ako možno ešte zefektívniť predaj aby boli tržby ešte vyššie, alebo vytvoriť novú prevádzku. To isté platí aj o ďalších mestách.

```
#vykonanie selectu v db pekaren a zapisanie do premennej
result = dbSendQuery(mysqlconnection,
"select sum(pp.predajna_cena) as celkove_trzby, dc.rok, dp.nazov_predajne,
dp.nazov_ulice, dp.mesto
from predaj_platba pp
left join d_cas dc on pp.id_cas = dc.id_cas
left join d_predajna dp on pp.id_predajna = dp.id_predajna
group by dc.rok, dp.id_predajna
order by celkove_trzby desc")
```

```
## Warning in .local(conn, statement, ...): Decimal MySQL column 0 imported as
## numeric
```

```
#vysledok dopytu
data.frame = fetch(result, n = 1000)
print(data.frame)
```

```
##    celkove_trzby  rok nazov_predajne          nazov_ulice            mesto
## 1        5199.41 2016         Bageta   Krajinská cesta 299        Svätý Jur
## 2        5158.34 2017         Bageta     Trnavská cesta 56       Bratislava
## 3        5133.10 2021         Bageta     Trnavská cesta 56       Bratislava
## 4        5132.81 2021         Bageta         Česká 1452/28          Galanta
## 5        5115.89 2018         Bageta   Krajinská cesta 299        Svätý Jur
## 6        5115.25 2019         Bageta    M. R. Štefánika 24          Pezinok
## 7        5106.91 2017         Bageta        Hlavná 50/1021          Stupava
## 8        5099.91 2020         Bageta          Halenárska 6           Trnava
## 9        5088.68 2016         Bageta        Košariská 3202   Dunajská Lužná
## 10       5043.35 2018         Bageta         Česká 1452/28          Galanta
```

```
## 11       5019.71 2016        Bageta        Hlavná 50/1021           Stupava
## 12       5004.46 2019        Bageta          Karloveská 5        Bratislava
## 13       4996.71 2016        Bageta          Karloveská 5        Bratislava
## 14       4994.43 2018        Bageta          Karloveská 5        Bratislava
## 15       4980.91 2017        Bageta       Kalinčiakova 4             Modra
## 16       4969.06 2018        Bageta       Košariská 3202    Dunajská Lužná
## 17       4961.45 2020        Bageta         Komenského 8 Ivanka pri Dunaji
## 18       4948.33 2021        Bageta   Krajinská cesta 299         Svätý Jur
## 19       4938.28 2018        Bageta     Trnavská cesta 56        Bratislava
## 20       4929.92 2018        Bageta  Hodžovo námestie 568        Bratislava
## 21       4928.14 2020        Bageta   Krajinská cesta 299         Svätý Jur
## 22       4907.20 2020        Bageta       Košariská 3202    Dunajská Lužná
## 23       4904.33 2019        Bageta       Kalinčiakova 4             Modra
## 24       4900.88 2020        Bageta        Česká 1452/28           Galanta
## 25       4898.07 2020        Bageta       Kalinčiakova 4             Modra
## 26       4888.71 2019        Bageta  Hodžovo námestie 568        Bratislava
## 27       4887.69 2018        Bageta Janka Jesenského 3653             Senec
## 28       4883.46 2020        Bageta     Trnavská cesta 56        Bratislava
## 29       4882.85 2021        Bageta  Hodžovo námestie 568        Bratislava
## 30       4882.11 2021        Bageta          Karloveská 5        Bratislava
## 31       4873.02 2017        Bageta          Karloveská 5        Bratislava
## 32       4871.99 2016        Bageta Janka Jesenského 3653             Senec
## 33       4860.40 2019        Bageta       Košariská 3202    Dunajská Lužná
## 34       4857.52 2016        Bageta          Slovenská 10           Malacky
## 35       4852.83 2020        Bageta    M. R. Štefánika 24           Pezinok
## 36       4847.53 2017        Bageta          Halenárska 6            Trnava
## 37       4837.47 2017        Bageta          Komenského 8 Ivanka pri Dunaji
## 38       4836.86 2016        Bageta  Hodžovo námestie 568        Bratislava
## 39       4825.52 2020        Bageta  Hodžovo námestie 568        Bratislava
## 40       4823.46 2020        Bageta        Hlavná 50/1021           Stupava
## 41       4822.59 2017        Bageta    M. R. Štefánika 24           Pezinok
## 42       4817.24 2019        Bageta   Krajinská cesta 299         Svätý Jur
## 43       4799.13 2017        Bageta         Česká 1452/28           Galanta
## 44       4789.63 2017        Bageta Janka Jesenského 3653             Senec
## 45       4788.32 2018        Bageta          Halenárska 6            Trnava
## 46       4754.27 2016        Bageta          Komenského 8 Ivanka pri Dunaji
## 47       4744.89 2016        Bageta     Trnavská cesta 56        Bratislava
## 48       4742.41 2019        Bageta          Komenského 8 Ivanka pri Dunaji
## 49       4727.84 2019        Bageta         Česká 1452/28           Galanta
## 50       4720.15 2019        Bageta     Trnavská cesta 56        Bratislava
## 51       4720.05 2021        Bageta          Halenárska 6            Trnava
## 52       4717.51 2021        Bageta          Komenského 8 Ivanka pri Dunaji
## 53       4716.06 2018        Bageta    M. R. Štefánika 24           Pezinok
## 54       4715.22 2019        Bageta        Hlavná 50/1021           Stupava
## 55       4711.02 2016        Bageta    M. R. Štefánika 24           Pezinok
## 56       4710.02 2017        Bageta  Hodžovo námestie 568        Bratislava
## 57       4705.37 2021        Bageta Janka Jesenského 3653             Senec
## 58       4701.90 2018        Bageta       Kalinčiakova 4             Modra
## 59       4699.03 2017        Bageta   Krajinská cesta 299         Svätý Jur
## 60       4698.49 2017        Bageta       Košariská 3202    Dunajská Lužná
## 61       4690.39 2021        Bageta    M. R. Štefánika 24           Pezinok
## 62       4678.50 2018        Bageta          Komenského 8 Ivanka pri Dunaji
## 63       4675.61 2018        Bageta          Slovenská 10           Malacky
## 64       4674.08 2018        Bageta        Hlavná 50/1021           Stupava
```

```
## 65       4663.35 2016         Bageta          Halenárska 6            Trnava
## 66       4657.41 2021         Bageta        Košariská 3202    Dunajská Lužná
## 67       4647.40 2021         Bageta          Slovenská 10           Malacky
## 68       4639.97 2016         Bageta        Kalinčiakova 4             Modra
## 69       4591.99 2020         Bageta          Karloveská 5        Bratislava
## 70       4581.76 2021         Bageta        Kalinčiakova 4             Modra
## 71       4548.61 2019         Bageta          Halenárska 6            Trnava
## 72       4540.62 2020         Bageta          Slovenská 10           Malacky
## 73       4533.93 2020         Bageta Janka Jesenského 3653             Senec
## 74       4515.37 2019         Bageta Janka Jesenského 3653             Senec
## 75       4483.63 2022         Bageta          Karloveská 5        Bratislava
## 76       4475.35 2021         Bageta        Hlavná 50/1021           Stupava
## 77       4369.90 2022         Bageta        Košariská 3202    Dunajská Lužná
## 78       4365.34 2022         Bageta        Hlavná 50/1021           Stupava
## 79       4350.85 2019         Bageta          Slovenská 10           Malacky
## 80       4321.88 2016         Bageta         Česká 1452/28           Galanta
## 81       4309.81 2022         Bageta    M. R. Štefánika 24           Pezinok
## 82       4292.19 2017         Bageta          Slovenská 10           Malacky
## 83       4274.99 2022         Bageta     Trnavská cesta 56        Bratislava
## 84       4272.23 2022         Bageta   Krajinská cesta 299         Svätý Jur
## 85       4259.54 2022         Bageta  Hodžovo námestie 568        Bratislava
## 86       4202.60 2022         Bageta         Česká 1452/28           Galanta
## 87       4177.57 2022         Bageta          Komenského 8 Ivanka pri Dunaji
## 88       4172.08 2022         Bageta Janka Jesenského 3653             Senec
## 89       4081.17 2022         Bageta          Slovenská 10           Malacky
## 90       4045.01 2022         Bageta        Kalinčiakova 4             Modra
## 91       3947.09 2022         Bageta          Halenárska 6            Trnava
```

Otázka č. 2 - Aké výrobky sa predávajú najviac v lete, a aké v zime?

Interpretácia: Za jednotlivé mesiace sú najpredávanejšie výrobky a ich počty zobrazené v tabuľke. Z dopytu vidíme, že prvých 6 výrobkov sa predáva najviac, mohli by sme na ne v daný mesiac poskytnúť napríklad množstevnú zľavu.

```
#vykonanie selectu v db pekaren a zapisanie do premennej
result = dbSendQuery(mysqlconnection,
"select COUNT(pp.id_vyrobok) as pocet, dv.nazov_vyrobku, dc.mesiac
from predaj_platba pp
left join d_cas dc on pp.id_cas = dc.id_cas
left join d_vyrobok dv on pp.id_vyrobok = dv.id_vyrobok
left join d_predajna dp on pp.id_predajna = dp.id_predajna
where dc.mesiac in (1,2,6,7,8,12)
group by dc.mesiac
")
#vysledok dopytu
data.frame = fetch(result, n = 1000)
#print(data.frame)
head(data.frame, 1000)
```

```
##   pocet                                    nazov_vyrobku mesiac
## 1 14192                          pastiersky chlieb bez E      1
## 2 12430                                  rožok grahamový     12
## 3 12806 bona vita active krehké rožteky s celozrnnou múkou      2
## 4 13770                         chlieb tradičný kváskový      6
## 5 14497                            chlieb ražno-špaldový      7
```

```
## 6 14490                                          pletenka      8
```

Otázka č. 1 - V akej predajni a v akom čase sa predáva tovar najviac ?

Interpretácia: V jednotlivých predajniach sa tovar predáva najviac v danom čase. Z tohto dopytu by sme mohli určiť napr. happy hour, ktorá by bola v danej hodine na danej predajni. Zákazník by na celý nákup získal zľavu vo výške danej hodiny.

```
#vykonanie selectu v db pekaren a zapisanie do premennej
result = dbSendQuery(mysqlconnection,
"SELECT hodina, COUNT(hodina) AS `pocet_vyskytov`, dp.nazov_ulice, dp.mesto
FROM predaj_platba pp
left join d_cas dc on dc.id_cas = pp.id_cas
left join d_predajna dp on dp.id_predajna = pp.id_predajna
GROUP BY hodina
ORDER BY `pocet_vyskytov` DESC
")
#vysledok dopytu
data.frame = fetch(result, n = 1000)
#print(data.frame)
head(data.frame, 1000)
```

```
##    hodina pocet_vyskytov          nazov_ulice          mesto
## 1      10          14094 Hodžovo námestie 568     Bratislava
## 2       7          14067  Krajinská cesta 299      Svätý Jur
## 3       9          13995    Trnavská cesta 56     Bratislava
## 4      15          13929       Kalinčiakova 4          Modra
## 5       6          13923 Hodžovo námestie 568     Bratislava
## 6       8          13921       Hlavná 50/1021        Stupava
## 7      11          13849  Krajinská cesta 299      Svätý Jur
## 8      12          13728         Halenárska 6         Trnava
## 9      17          13707  Krajinská cesta 299      Svätý Jur
## 10     13          13551    Trnavská cesta 56     Bratislava
## 11     16          13474       Kalinčiakova 4          Modra
## 12     14          13405       Košariská 3202 Dunajská Lužná
```