ANALYSIS ON HOTEL REVIEW DATA

MINIJA KANNATH

CONTENT

- PROBLEM DEFINITION
- DATA COLLECTION
- ER DIAGRAM AND SCHEMA
- DATA PREPARATION
- DATA ANALYSIS
- DATA VISUALIZATION AND INSIGHTS

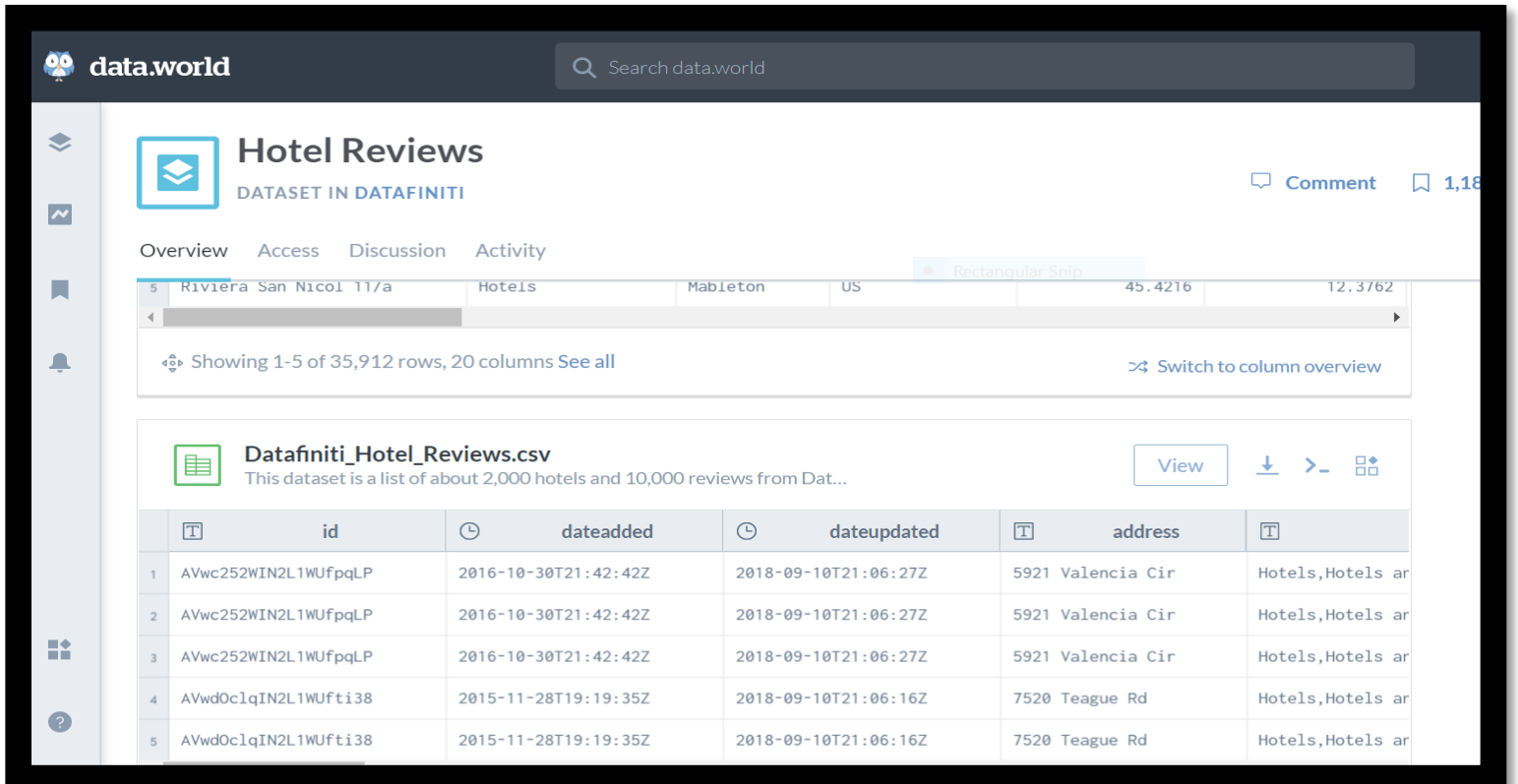
PROBLEM DEFINITION

The objective of this analysis is to find out which place in LA is more profitable to start a new hotel. In this analysis a comparative study has been done on low rated hotels and high rated hotels in LA. Customer feedback in terms of ratings and reviews collected from hotels located in LA has been analysed .The fundamental issue for the survival and growth of hotels is measurement and management of service quality and customer experience.

DATA COLLECTION

This data is collected from Data.world.

This is a list of 2,000 hotels in US and their reviews provided by Datafiniti's Business Database. The dataset includes hotel location, name, rating, review data, title, username, and more.



THE RAW DATA

id	dateAdde	address	categories	primaryCa	city	country	latitude	longitude	name	province	reviews.d	reviews.d	reviews.ra	reviews.tit	reviews.userCity	reviews.username
AVwc252V	2016-10-3	5921 Vale	Hotels,Ho	Accommo	Rancho Sa	US	32.99096	-117.186	Rancho Va	CA	2013-11-1	2016-08-0	5	Best romantic vacation ever!!!!		Paula
AVwc252V	2016-10-3	5921 Vale	Hotels,Ho	Accommo	Rancho Sa	US	32.99096	-117.186	Rancho Va	CA	2014-07-0	2016-08-0	5	Sweet sweet serenity		D
AVwc252V	2016-10-3	5921 Vale	Hotels,Ho	Accommo	Rancho Sa	US	32.99096	-117.186	Rancho Va	CA	2015-01-0	2016-11-1	5	Amazing Property and Experien		Ron
AVwdOclq	2015-11-2	7520 Teag	Hotels,Ho	Accommo	Hanover	US	39.15593	-76.7163	Aloft Arun	MD	2016-05-1	2016-05-2	2	Never aga	Richmond	jaeem2016
AVwdOclq	2015-11-2	7520 Teag	Hotels,Ho	Accommo	Hanover	US	39.15593	-76.7163	Aloft Arun	MD	2016-07-0	2016-07-3	5	ALWAYS G	Laurel	MamaNiaOne
AVwdOclq	2015-11-2	7520 Teag	Hotels,Ho	Accommo	Hanover	US	39.15593	-76.7163	Aloft Arun	MD	2016-06-1	2016-07-3	5	Wonderfu	Laurel	kevan777
AVwdOclq	2015-11-2	7520 Teag	Hotels,Ho	Accommo	Hanover	US	39.15593	-76.7163	Aloft Arun	MD	2016-04-3	2016-05-0	5	Worth the money		Princess F
AVwdOclq	2015-11-2	7520 Teag	Hotels,Ho	Accommo	Hanover	US	39.15593	-76.7163	Aloft Arun	MD	2016-06-2	2016-07-1	5	Great Hot	Clayton	DebMurphy57
AVwdOclq	2015-11-2	7520 Teag	Hotels,Ho	Accommo	Hanover	US	39.15593	-76.7163	Aloft Arun	MD	2016-05-2	2016-07-3	5	Short stay	Boston	kayleighwillswim1224
AVwePiAX	2016-03-2	315 SE Oly	Hotels,Ho	Accommo	Vancouve	US	45.61921	-122.525	Hampton	WA	2016-01-2	2016-04-1	5	Amazing e	Portland	KristyWM
AVwePiAX	2016-03-2	315 SE Oly	Hotels,Ho	Accommo	Vancouve	US	45.61921	-122.525	Hampton	WA	2016-05-0	2016-05-1	5	I loved our stay here		B M
AVwePiAX	2016-03-2	315 SE Oly	Hotels,Ho	Accommo	Vancouve	US	45.61921	-122.525	Hampton	WA	2016-01-3	2016-04-1	5	Hampton	Antioch	Cathleen S
AVwePiAX	2016-03-2	315 SE Oly	Hotels,Ho	Accommo	Vancouve	US	45.61921	-122.525	Hampton	WA	2016-03-1	2016-03-2	5	Perfection	Portland	1fiesty
AVwePiAX	2016-03-2	315 SE Oly	Hotels,Ho	Accommo	Vancouve	US	45.61921	-122.525	Hampton	WA	2016-06-2	2016-07-2	5	Good hote	Port Orchard	810michellemm
AVwePiAX	2016-03-2	315 SE Oly	Hotels,Ho	Accommo	Vancouve	US	45.61921	-122.525	Hampton	WA	2016-06-2	2016-07-2	5	Excellent!	San Diego	travelchick1113
AVwdUKE	2016-11-0	106 W 12t	Hotels,Cat	Accommo	Kansas Cit	US	39.10012	-94.5847	Hotel Phill	MO	2015-09-1	2016-03-2	4	Very pleased		AimlessinMe
AVwdUKE	2016-11-0	106 W 12t	Hotels,Cat	Accommo	Kansas Cit	US	39.10012	-94.5847	Hotel Phill	MO	2015-11-2	2016-03-2	5	Excellent Stay		Tabitha K
AVwdUKE	2016-11-0	106 W 12t	Hotels,Cat	Accommo	Kansas Cit	US	39.10012	-94.5847	Hotel Phill	MO	2016-01-3	2016-03-2	3	Beautiful lobby but rooms neec		Yamachaser
AVwdUKE	2016-11-0	106 W 12t	Hotels,Cat	Accommo	Kansas Cit	US	39.10012	-94.5847	Hotel Phill	MO	2015-12-1	2016-03-2	5	Hotel's 19	Wichita	Rbuffington
AVwdUKE	2016-11-0	106 W 12t	Hotels,Cat	Accommo	Kansas Cit	US	39.10012	-94.5847	Hotel Phill	MO	2015-09-0	2016-03-2	4	Excellent s	Chicago	Nick F

E-R DIAGRAM

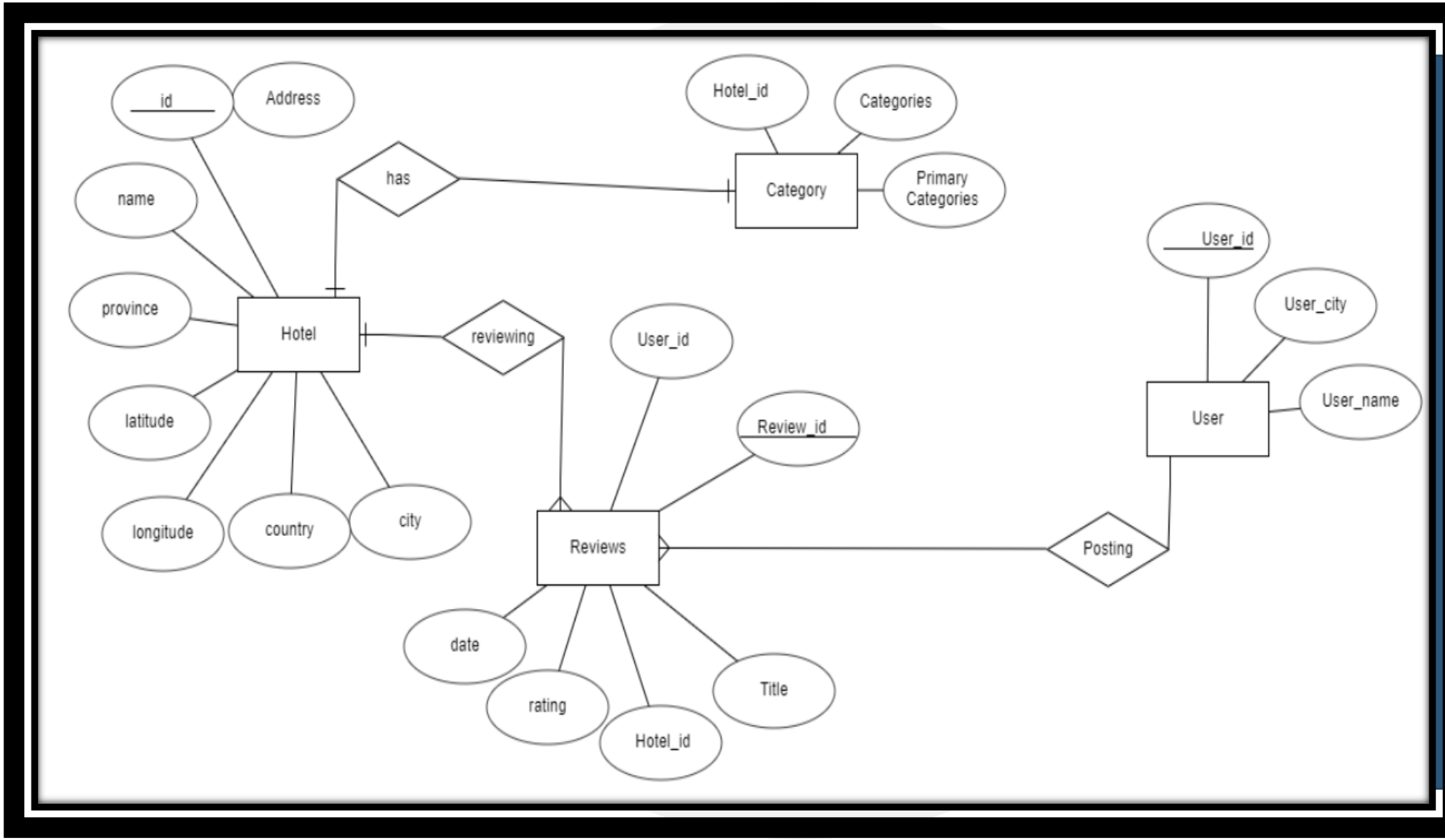
Entity

1.Review

2.Hotel

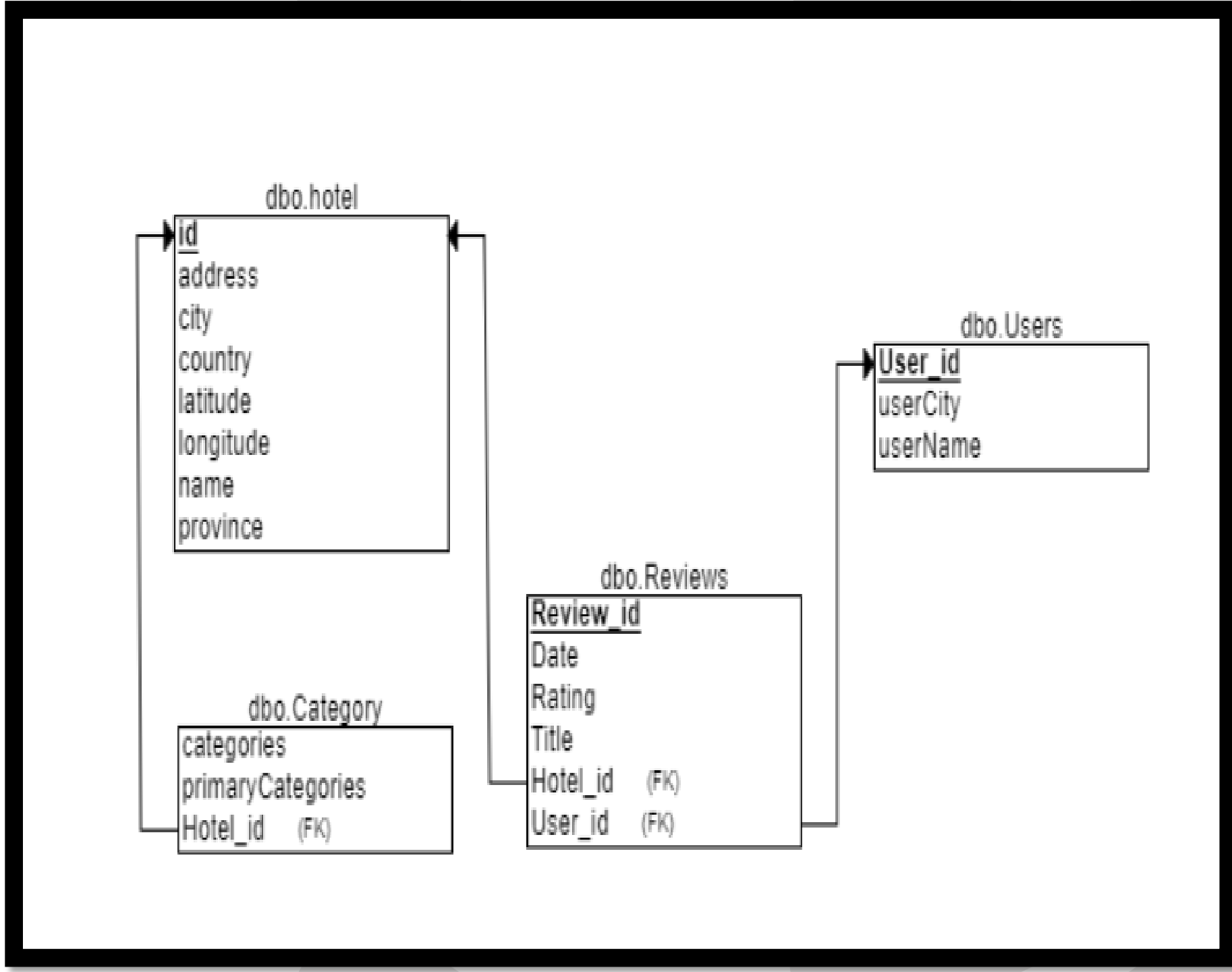
3.Category

4.User

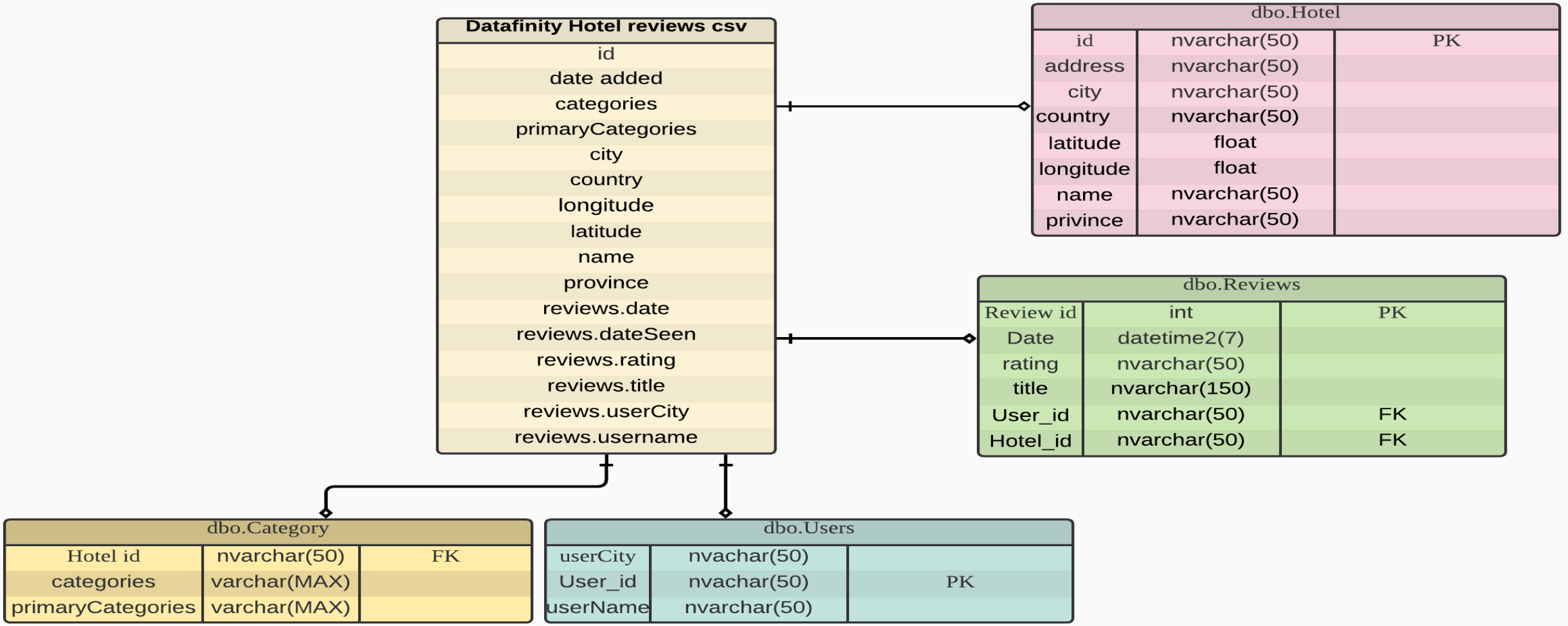


Data base diagram

FK –Foreign key



CREATING TABLES BASED ON SCHEMA



Data Analysis

To start with data Analysis Tables are imported into the sql server .Primary key and foreign key are assigned.

```
ALTER TABLE [dbo].[Hotel]
ADD PRIMARY KEY (id);
ALTER TABLE [dbo].[Reviews]
ADD PRIMARY KEY (Review_id);
ALTER TABLE [dbo].[Users]
ADD PRIMARY KEY (User_id);
--to assaign foreign key
ALTER TABLE [dbo].[Reviews]

ADD FOREIGN KEY (Hotel_id) REFERENCES [dbo].[Hotel](id);
ALTER TABLE [dbo].[Reviews]
```

```
--to assaign foreign key
ALTER TABLE [dbo].[Reviews]

ADD FOREIGN KEY (Hotel_id) REFERENCES [dbo].[Hotel](id);
ALTER TABLE [dbo].[Reviews]

ADD FOREIGN KEY (User_id) REFERENCES [dbo].[Users](User_id);
```

DATA ANALYSIS

To find out top and low rated hotels in LA

```
with Hotel as
(select   h.name,c.categories,round(avg(r.Rating),1) as 'stars',h.province
from [dbo].[Hotel] h inner join  [dbo].[Reviews] r
on h.id = r.Hotel_id
inner join [dbo].[Category]c
on h.id = c.Hotel_id
group by c.categories,h.name,h.province
having h.province = 'LA')

select name,stars,
case
    when stars between 1.0 and 3.0 then 'low rated'
    when stars between 3.1 and 5 then 'high rated'
    end rating
from Hotel;
```

name	stars	rating
Ace Hotel New Orleans	3.3	high rated
Aloft New Orleans Downtown	4.2	high rated
Baton Rouge Marriott	3	low rated
Bendel Executive Suites	5	high rated
Best Western Airport Inn	4	high rated
Best Western Casino Inn	3	low rated
Best Western Of Alexandria Inn & Suites & Conferen...	3	low rated
Best Western Plus Chalmette Hotel	5	high rated
Best Western Plus Deridder Inn & Suites	4	high rated
Best Western Plus Executive Hotel Suites	4.5	high rated
Best Western Plus Richmond Inn & Suites-Baton Ro...	4	high rated
Econo Lodge	1	low rated
Econo Lodge	2	low rated
Empress Hotel	1.7	low rated
French Quarter Suites Hotel	3.6	high rated
Hampton Inn Suites Monroe	5	high rated
Hampton Inn Sulphur/Lake Charles Area	5	high rated
Hilton Garden Inn Shreveport Bossier City	5	high rated
Holiday Inn Express & Suites Baton Rouge East	5	high rated
Hotel Royal	4.5	high rated

DATA ANALYSIS

To find out the 3 top Cities with most number of hotels in LA

```
with city as(
select province,city,longitude,latitude
from [dbo].[Hotel]

group by city,province,name,longitude,latitude
Having province ='LA')
select  Top 3 count(city) as 'citycount',city
from city
group by city
order by count(city) desc;
```

citycount	city
18	New Orleans
5	Lafayette
3	Baton Rouge

DATA ANALYSIS

To find out the distances of hotels in LA from Air port and city centre from latitude and longitude of the cities the Haversine Formula is used https://www.vcalc.com/wiki/vCalc/Haversine+-+Distance

```
--To find the longtitude and latitude

Select  h.name,round(avg(r.rating),2) as stars,h.latitude,h.longitude
from [dbo].[Hotel] h inner join [dbo].[Reviews] r
on h.id = r.Hotel_id
group by h.name,h.province,h.latitude,h.longitude
having h.province = 'LA'
order by round(avg(r.rating),2) desc;

select name,city
from [dbo].[Hotel];
```

name	stars	latitude	longitude
Bendel Executive Suites	5	30.20076	-92.0182
Best Western Plus Chalmette Hotel	5	29.950949	-89.960338
Best Western Plus Vermilion River Suites Hotel	5	30.19098	-92.01345
Bienville House	5	29.9538338	-90.06527
Hampton Inn Suites Monroe	5	32.49287	-92.05618
Hampton Inn Sulphur/Lake Charles Area	5	30.216843	-93.37432
Hilton Garden Inn Shreveport Bossier City	5	32.51738	-93.71154
Holiday Inn Express & Suites Baton Rouge East	5	30.393513	-91.06225
Hotel Storyville	5	29.96762	-90.06539
Mainstay Suites	5	29.56927	-90.68337
Maison de Macarty	5	29.96348	-90.0374
MOXY New Orleans Downtown/French Quarter Area	5	29.953565	-90.07317

DATA ANALYSIS

To find out the user who gives maximum reviews with average rating

```
--user with max reviews and below average ratings
with LowReview as(
    select u.userName,
    case
        when r.Rating between 1 and 2 Then 'BelowAverage'
    End Rating
    from [dbo].[Reviews] r inner join [dbo].[Users] u
    on r.User_id = u.User_id)

select Top 3 userName,count(Rating) as 'Star_count'
from LowReview
group by Rating,userName
order by count(Rating) desc;
```

userName	Star_count
A verified traveler	153
A Traveler	89
John	6

DATA ANALYSIS

To find out why hotels are rated bad and good from reviewer's text

```
--To find the good features

with Hotel as(
select   h.name,round(avg(r.Rating),1) as 'stars',h.province,r.Title
from [dbo].[Hotel] h inner join  [dbo].[Reviews] r
on h.id = r.Hotel_id

group by h.name,h.province,r.Title
having h.province = 'LA')

select
    CASE
        when upper(Title) like '%Stay%' then 'Good stay'
  when upper(Title) like '% Loc%' then 'Location good'
  when upper(Title) like '%Awesome%' then 'Awesome hotel'
  when upper(Title) like '%Service%' then 'Good service'
  when upper(Title) like '%Best%' then 'Best hotel'
  when upper(Title) like '%Class%' then 'Classic'
  when upper(Title) like '%Excellent%' then 'Excellent Hotel'
  when upper(Title) like '%accoma%' then 'Good accomodation'
  when upper(Title) like '%F%' then 'Friendly staff'
  when upper(Title) like '%Great%' then 'Great Hotel'
        when upper(Title) like '%Love%' then 'Love this Hotel'
        when upper(Title) like '%Laxu%' then 'Luxury comfort'
        when upper(Title) like '%perfect%' then 'perfect place'
        when upper(Title) like '%Very nice%' then 'Great Hotel'
        when upper(Title) like '%Wonder%' then 'Woderful experience'
        else 'other'

    END Good_Title
    ,count(*) as count
    from Hotel
    group by Title;
```

Good_Title	count
Friendly staff	1
Good stay	1
Great Hotel	1
other	1
other	1
other	1
Friendly staff	1
Friendly staff	1
other	1
other	1
Friendly staff	1
other	1

```
---bad reviews

    with Hotel as(
select   h.name,round(avg(r.Rating),1) as 'stars',h.province,r.Title
from [dbo].[Hotel] h inner join  [dbo].[Reviews] r
on h.id = r.Hotel_id

group by h.name,h.province,r.Title
having h.province = 'LA')


select Title,
    CASE
  when upper(Title) like '%Soiled%' then 'Soiled Bed'
  when upper(Title) like '% not%' then 'Do not go there'
  when upper(Title) like '%Disappointing%' then 'Disappointing Experience'
  when upper(Title) like '%Service%' then 'Bad service'
  when upper(Title) like '%Food%' then 'Disappointing Food'
  when upper(Title) like '%location%' then 'Bad location'
  when upper(Title) like '%internet%' then 'No wifi'
  when upper(Title) like '%Average%' then 'Average Hotel'
  when upper(Title) like '%W%' then 'Not worth money'
  when upper(Title) like '%s%' then 'Small rooms'
  else 'other'

  END Bad_Title
  ,count(*) as count
  from Hotel
  group by Title;
```

	Bad_Title	count
nd ...	Soiled Bed	1
	Small rooms	1
	other	1
	other	1
	Not worth money	1
	Small rooms	1
	Small rooms	1
	Small rooms	1
y	Not worth money	1
	Not worth money	1
	other	1
	Not worth money	1

DATA VISUALIZATION

To start with the visualization query results are imported into excel and excel charts are used to create charts.

Questions analyzed:

- Is there a correlation between the distance from airport and rating of the hotel .
- Is there a correlation between the distance from city center and Rating of the hotel.
- The city with maximum number of hotels.
- The features affecting the rating and reviews
- Top reviewer with average rating.

CONCLUSION

This data exploration and visualisation helped us to gather a few useful insights about the Hotels in LA.Based on this Analysis we found out that the following features have contributed in getting good reviews and high ratings.

- Location
- Sleep quality
- Service
- Value

So we see that room and services are important variables for consideration. The major keywords from these hotels that were specially emphasized on were Hotel, Room, Staff, Great, Good, Services, Location. All these should be given importance when starting a new hotel.

THANK YOU

REFERENCES:HTTPS://DATA.WORLD/DATAFINITI/HOTEL-REVIEWS/WORKSPACE/FILE?FILENAME=DATAFINITI_HOTEL_REVIEWS.CSV