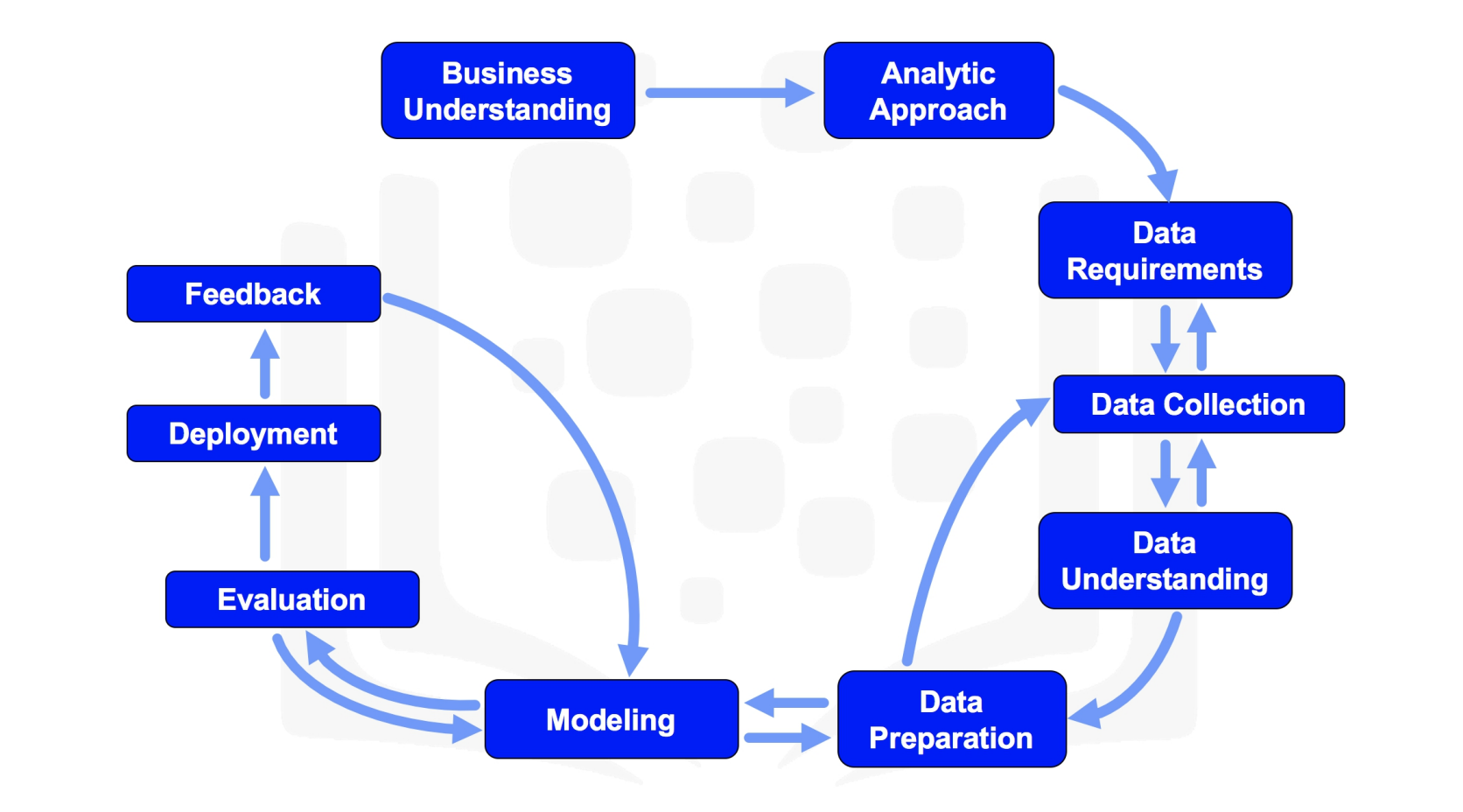
Choosing Neighborhood

IBM Data Science Capstone Project

Umut Aktaş

Data Science Methodology Steps



In this project, I followed the above schema on every step.

Description of the Problem
Business Understanding

- Our customers want to move from Moscow to Berlin.
- They want to choose the district in which they will find same services that they are used to have in Moscow.
- They need a gym in the neighborhood, some nice cafe, bars and restaurants.
- Now they are living in Moscow Khamovniki and it is a very good district. We are searching for district in Berlin that will have same venues on the same distances from the center of the district.

Short Recap of Moscow Khamovniki District

The district extends from Bolshoy Kamenny Bridge into the Luzhniki bend of Moskva River; northern boundary with Arbat District follows Znamenka Street, Gogolevsky Boulevard, Sivtsev Vrazhek and Borodinsky Bridge.

The district contains Pushkin Museum, Cathedral of Christ the Saviour, Devichye Pole medical campus, Novodevichy Convent and memorial cemetery, Luzhniki Stadium. The stretch of Khamovniki between Boulevard Ring and Garden Ring, known as Golden Mile, is downtown Moscow's most expensive housing area.

Short Recap of Berlin

Berlin is the capital and largest city of Germany by both area and population.

Berlin straddles the banks of the River Spree, which flows into the River Havel (a tributary of the River Elbe) in the western borough of Spandau. Among the city's main topographical features are the many lakes in the western and southeastern boroughs formed by the Spree, Havel, and Dahme rivers (the largest of which is Lake Müggelsee)

Berlin’s boroughs include old industrial areas that have become thriving cultural centres, as well as sleepy villages, busy markets and arcades, and endless green landscapes, rivers and bathing lakes. But above all there are **12 districts and 95 neighborhoods with average that we can suggest to our customer**.

Discussion of the background
Analytic Approach

- I wanted to focus ratings, house priceses, venues categories and their distances but unfortunately foursquare api doesn't allow us to take enough amount of rating data. So I focused only distance and categories of neighborhoods.
- With this data science project I focused center locations of neighborhoods, because I don't know our customer choosed house in Berlin yet.
- Of course, all notebook is iterable so you can change the address and run all code!

Description of the Data

Data Collection-Understanding

- I used the power of foursquare API for taking data as JSON files.
- With the help of the Pandas library, I turned these files into pandas data frames.

These data frames included;

- Venues types
- Distances of venues
- Longitudes and latitudes
- Names
- Neighborhoods
- Boroughs
- Addresses

Methodology

01 Data Preparation

In this stage I used power of foursquare api for taking data then turning to dataframes.
This case, I wanted to focus on ratings, house prices, venues categories, and their distances but unfortunately, the foursquare API doesn't allow us to take enough amount of rating data. So I focused only on distance and categories of neighborhoods.

- Restaurant
- Cafe
- Bar
- Gym

	name	categories	lat	lng	distance
0	Surf Coffee	Cafe	55.724556	37.562386	225
1	Pinzeria by Bontempi	Restaurant	55.727604	37.567514	390
2	Coffeemania (Кофемания)	Cafe	55.727374	37.567698	394
3	Анатомия. Персональный фитнес	Gym	55.722504	37.560276	455
4	Cosmoryba	Restaurant	55.727587	37.567575	393
5	Dance Studio Feelings 2	Gym	55.725440	37.562294	129
6	Cultura Coffee	Cafe	55.723236	37.560804	369
7	Double B Coffee & Tea (Даблби)	Cafe	55.728926	37.558408	332
8	The Truffle	Restaurant	55.727625	37.567455	387
9	Сыроварня	Restaurant	55.727667	37.567341	382
10	У Пиросмани	Restaurant	55.729064	37.558267	350
11	PRIME	Cafe	55.724217	37.564354	310
12	Крошка Картошка	Restaurant	55.727110	37.560834	79
13	Вино и Еда by Первый Нос	Bar	55.727229	37.567470	377
14	кофейня Буно	Cafe	55.725102	37.563493	198
15	Оджахури	Restaurant	55.727565	37.567694	399
16	Kakadu	Bar	55.727599	37.567623	396
17	Порт Находка	Restaurant	55.727585	37.567572	393

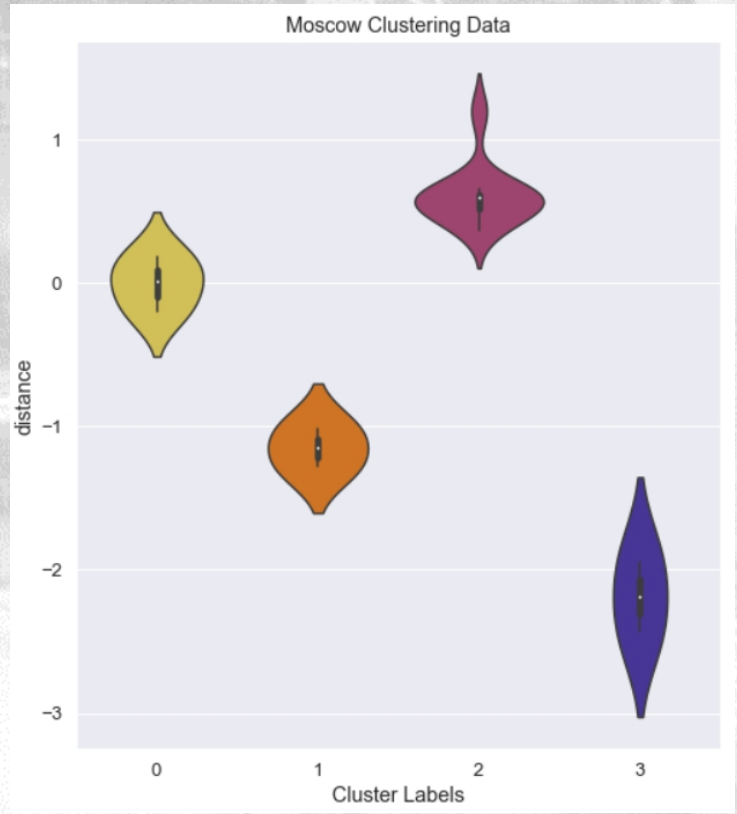
02 Feature Engineering

- I created new features for clarifying weights of this neighborhood.
- I summoned how many types of categories and mean distance of every each of them.

	categories	distance
0	Bar	1.258905
1	Cafe	-0.647042
2	Gym	-0.942056
3	Restaurant	0.330193

03 Modelling

- I used K-Means clustering considering with their distances.
- Clustering is a perfect way to comparing ths kid of data sets.

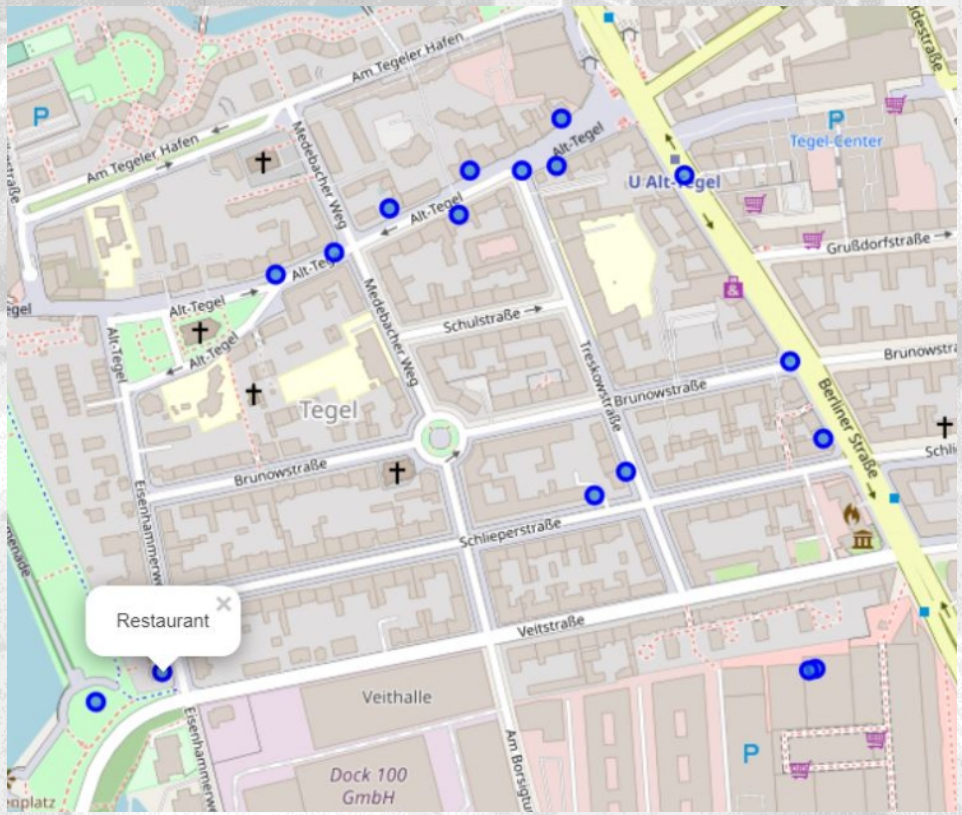


03 Modelling

I searched same neighborhoods in Berlin neighborhoods with the help of K-Means clustering.

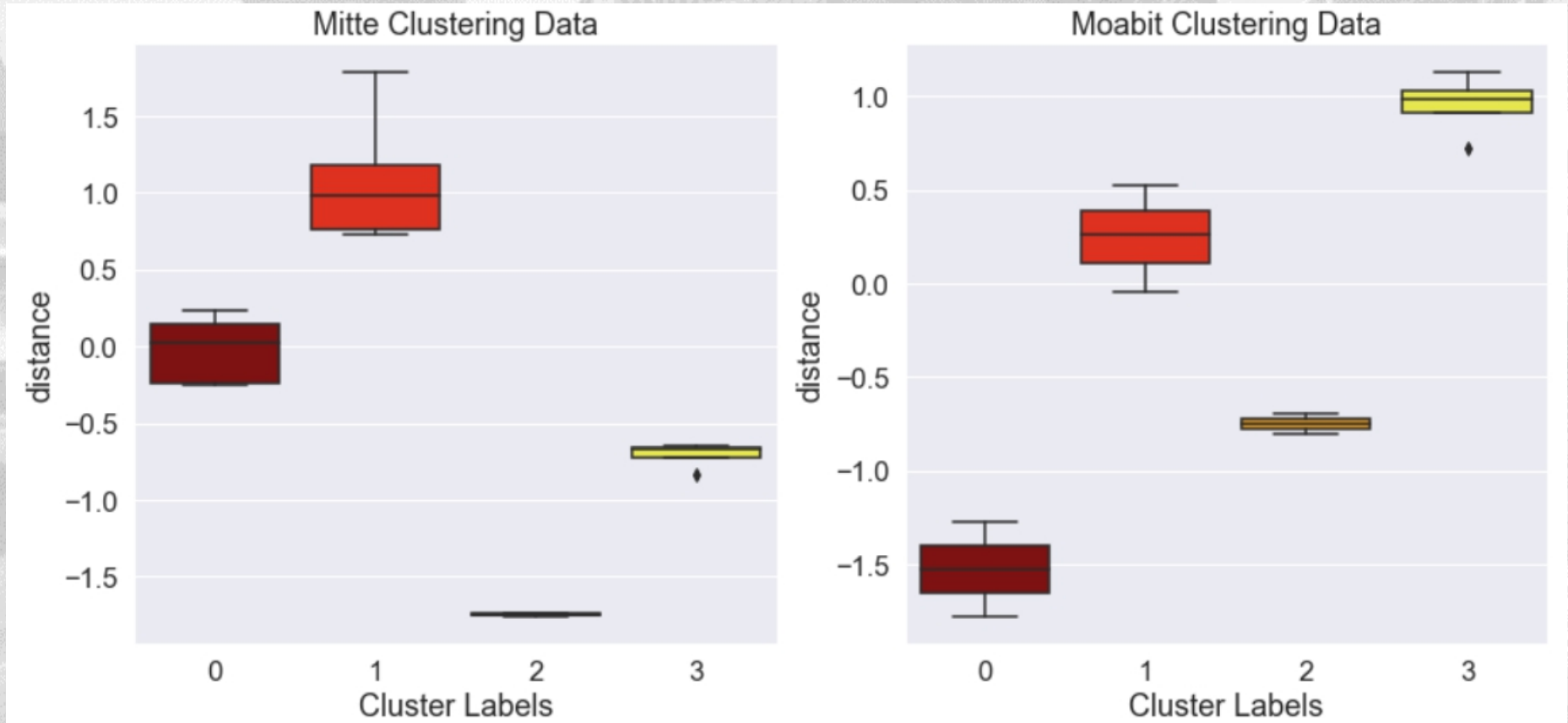
- Collecting latitudes and longtitudes of each neighborhood of Berlin
- Collecting json files for each neighborhoods
- Collecting weights of neighborhoods of Berlin
- Comparing with Moscow Khamovniki District
- Finding minimum mean distance values of neighborhood

```
1 Moabit has 4 same venues
2 Hansaviertel has 4 same venues
5 Gesundbrunnen has 4 same venues
43 Schöneberg has 4 same venues
63 Köpenick has 4 same venues
82 Fennpfuhl has 4 same venues
83 Rummelsburg has 4 same venues
85 Tegel has 4 same venues
```



04 Evaluation

In this section, I observed possible neighborhoods and how they match with our customer's neighborhood with K-Means clustering.



05 Result - Conclusion

- All of neighborhoods where are in last page have enough venue types and counts!
- With the help of observing our final illustration of K-means clustered plots and minimum mean distance value, I want to focus on which neighborhood has closer venues to the neighborhood center. Because in our case, I want to recommend the closest venues to our customer.
- With this data science project I focused center locations of neighborhoods, because I don't know our customer choosed house in Berlin yet.
- Mitte clusters have minimum distances. I can definitely recommend Mitte!
- All notebook is iterable so you can change Moscow Khamovniki to whereever you want!

Thank you for your time

LOREM IPSUM DOLOR LOREM

Result - Conclusion

- All of these neighborhoods have enough venue types and counts!
- With the help of observing our final illustration of K-means clustered plots and minimum mean distance value, I want to focus on which neighborhood has closer venues to the neighborhood center. Because in our case, I want to recommend the closest venues to our customer.
- With this data science project I focused center locations of neighborhoods, because I don't know our customer choosed house in Berlin yet.
- Mitte clusters have minimum distances. I can definitely recommend Mitte!
- !!! All notebook is iterable so you can change Moscow Khamovniki to whereever you want!

LOREM IPSUM DOLOR LOREM

LOREM IPSUM DOLOR

Lorem ipsum dolor sit am et, consectetur adipisicin g elit, sed do eiusmod te mpor incididunt ut labore et dolore magna aliqua. U t enim ad minim veniam,

Lorem ipsum dolor sit am et, consectetur adipisicin g elit, sed do eiusmod te mpor incididunt ut labore et dolore magna aliqua. U t enim ad minim veniam,

LOREM IPSUM DOLOR LOREM

LOREM IPSUM DOLOR

Lorem ipsum dol or sit amet, cons ectetur adipisicin g elit.

Lorem ipsum dol or sit amet, cons ectetur adipisicin g elit.

Lorem ipsum dol or sit amet, cons ectetur adipisicin g elit.

LOREM IPSUM DOLOR LOREM

LOREM IPSUM DOLOR

Lorem ipsum dolor sit amet, consectetur adipisicing elit.

Lorem ipsum dolor sit amet, consectetur adipisicing elit.

Lorem ipsum dolor sit amet, consectetur adipisicing elit.

Lorem ipsum dolor sit amet, consectetur adipisicing elit.

LOREM IPSUM DOLOR LOREM

LOREM IPSUM DOLOR

Lorem ipsum dolor sit amet, consectetur adipisicing elit.

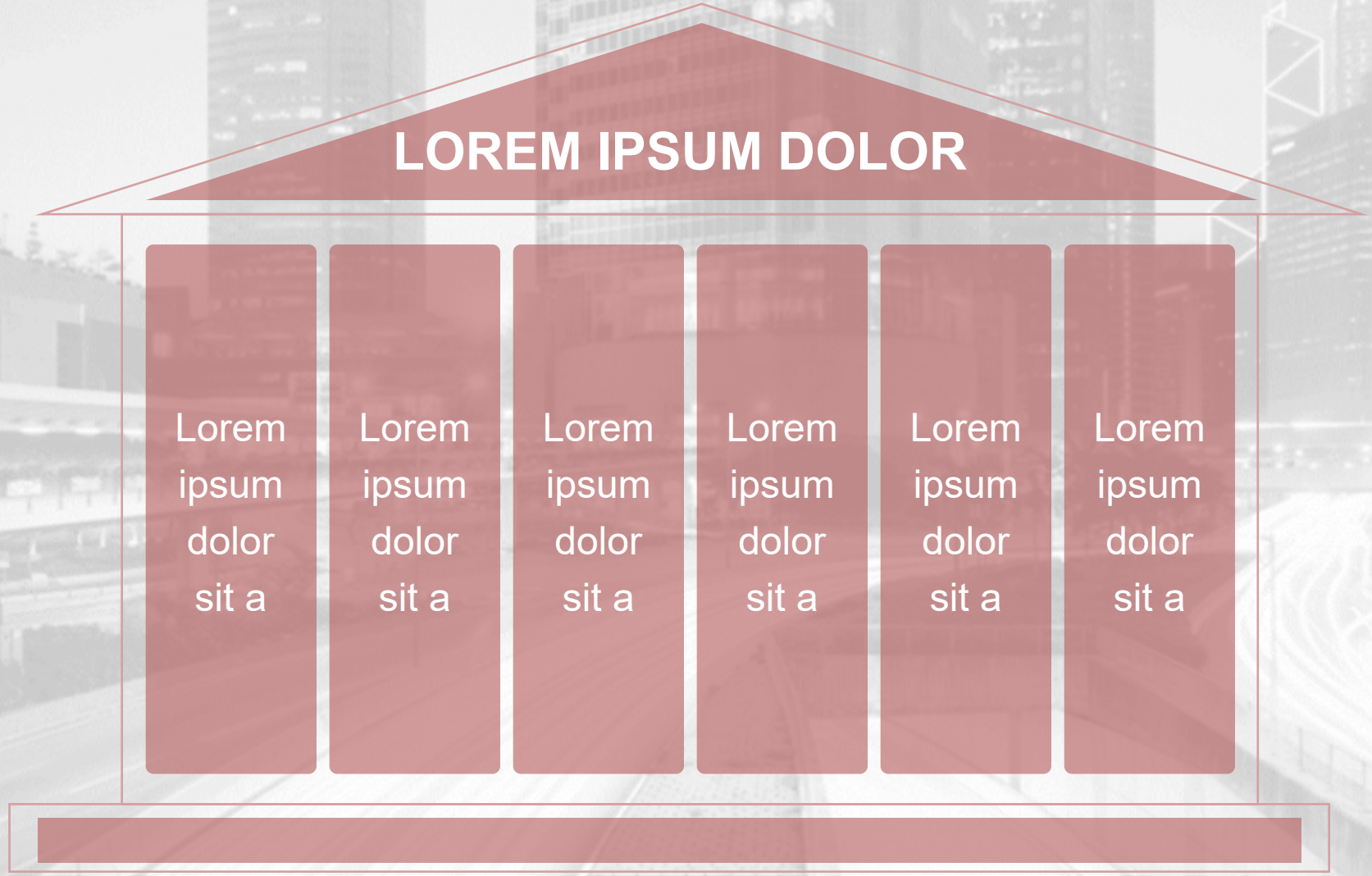
Lorem ipsum dolor sit amet, consectetur adipisicing elit.

Lorem ipsum dolor sit amet, consectetur adipisicing elit.

Lorem ipsum dolor sit amet, consectetur adipisicing elit.

Lorem ipsum dolor sit amet, consectetur adipisicing elit.

LOREM IPSUM DOLOR LOREM



LOREM IPSUM DOLOR LOREM

Lorem ipsum dolor sit amet, consectetur adipi sicing elit, sed do eiusmod tempor incididunt u t labore et dolore magna aliqua. Ut enim ad mi nim veniam, quis nostrud exercitation ullamco laboris nisi ut

LOREM IPSUM DOLOR LOREM

Lorem ipsum dolor sit amet, consectetur adipi sicing elit.

Lorem ipsum dolor sit amet, consectetur adipi sicing elit.

LOREM IPSUM DOLOR LOREM

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

谢谢大家！