



IBM CAPSTONE PROJECT – THE BATTLE OF NEIGHBORHOODS

Cluster Analysis of London Real Estate Market

A low-angle, black and white photograph of several tall skyscrapers reaching towards a cloudy sky. The perspective is from the ground looking up, creating a sense of height and scale. The buildings are modern with many windows. A large, semi-transparent white rectangle is overlaid on the left side of the image, containing the text. A green rectangular area is also visible, partially overlapping the white rectangle and the buildings.

BUSINESS PROBLEM SECTION

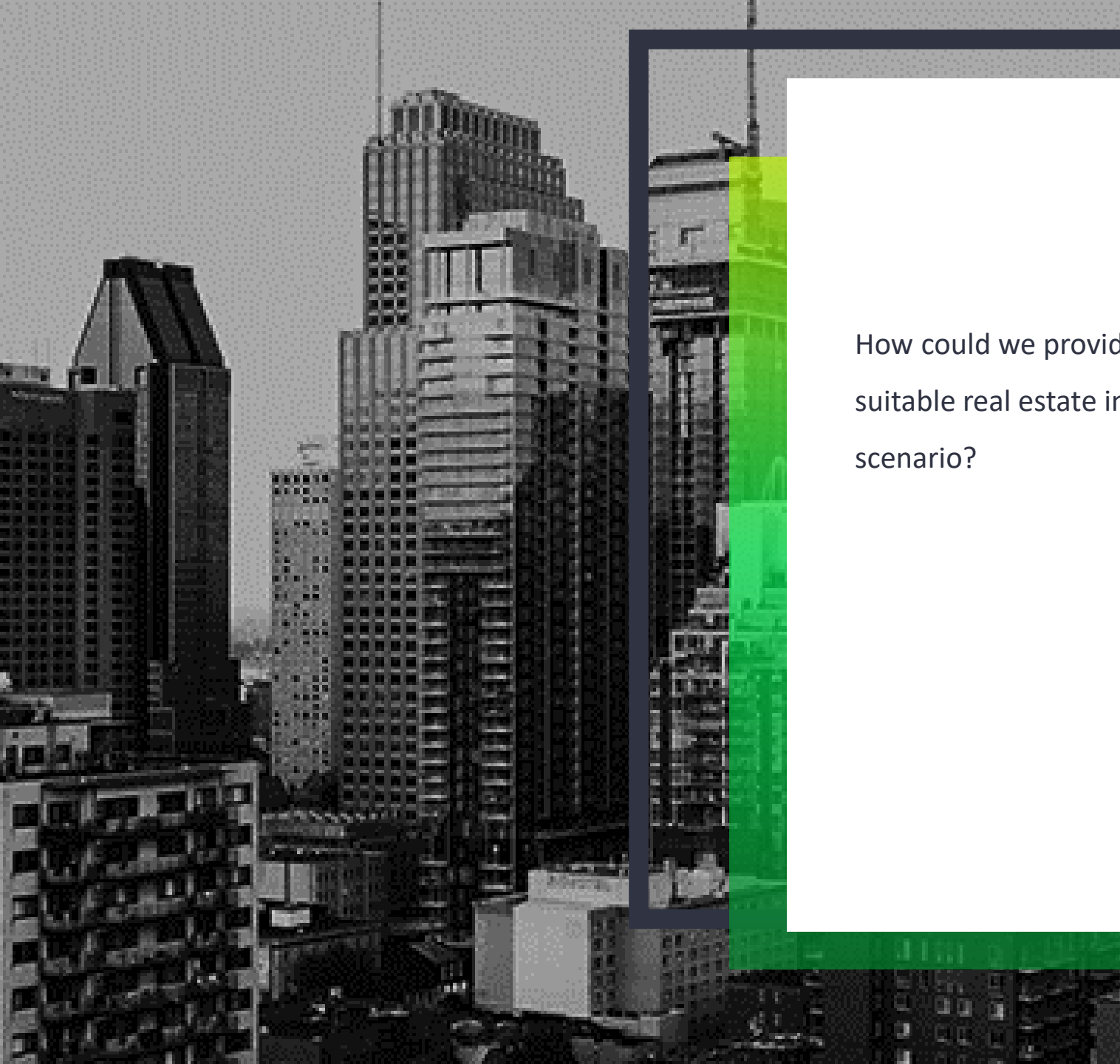
London Housing Market is in a rut:

1. Brexit
2. Hidden price falls
3. Record-low sales
4. Homebuilder exodus
5. Tax hikes addressing overseas buyers of homes in England and Wales.



The background is a grayscale aerial photograph of a dense urban area with numerous skyscrapers. Overlaid on this is a graphic consisting of a green rectangle with a white square inside it. A thick black L-shaped line frames the top-left corner of the white square. The text 'BUSINESS PROBLEM' is centered within the white square in a bold, black, sans-serif font.

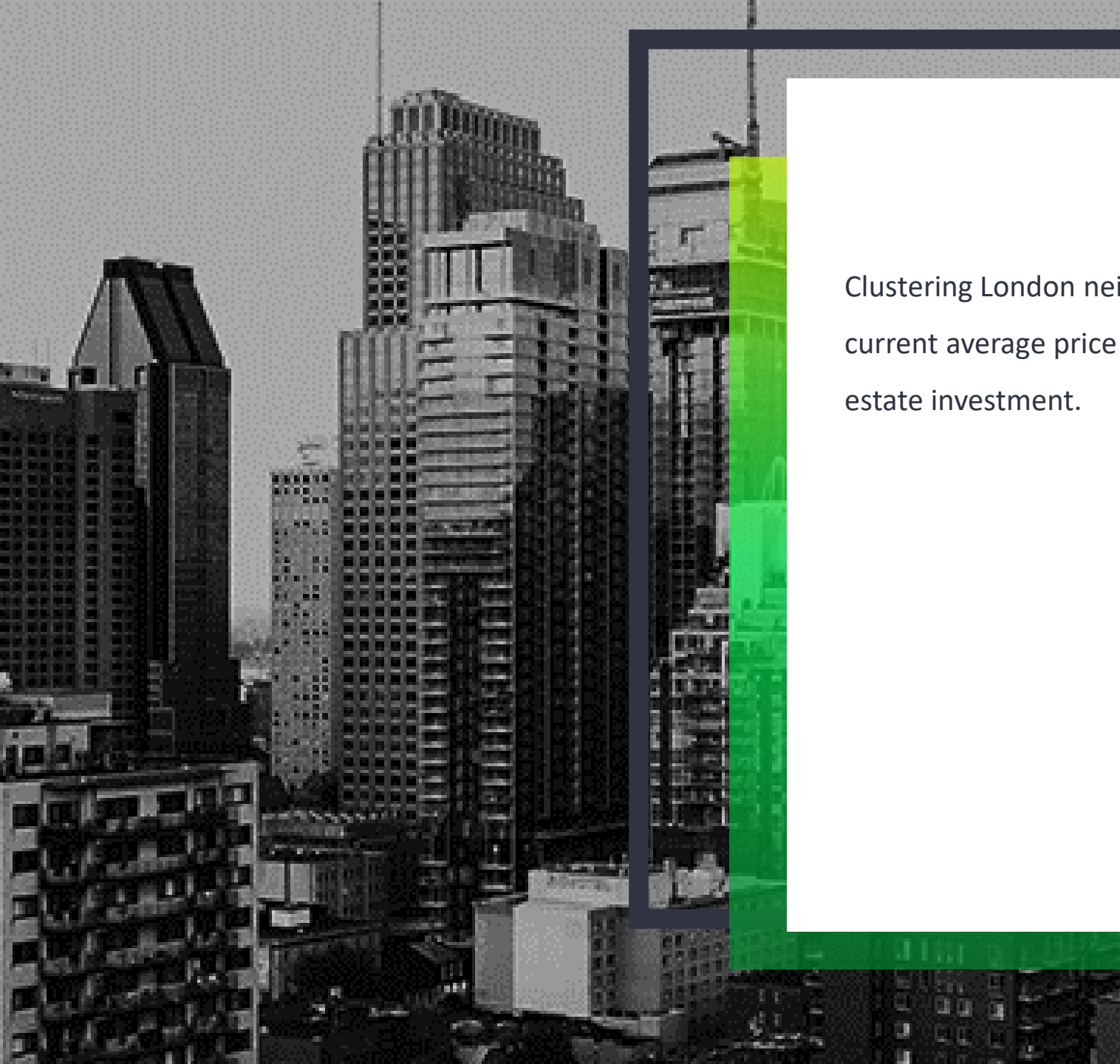
BUSINESS PROBLEM



How could we provide support to homebuyers clientele in to purchase a suitable real estate in London in this uncertain economic and financial scenario?

The background is a grayscale aerial photograph of a dense urban area with numerous skyscrapers. Overlaid on this is a graphic consisting of a green rectangle on the left and a white rectangle on the right, separated by a thick black L-shaped line. The word "SOLUTION" is centered in the white rectangle.

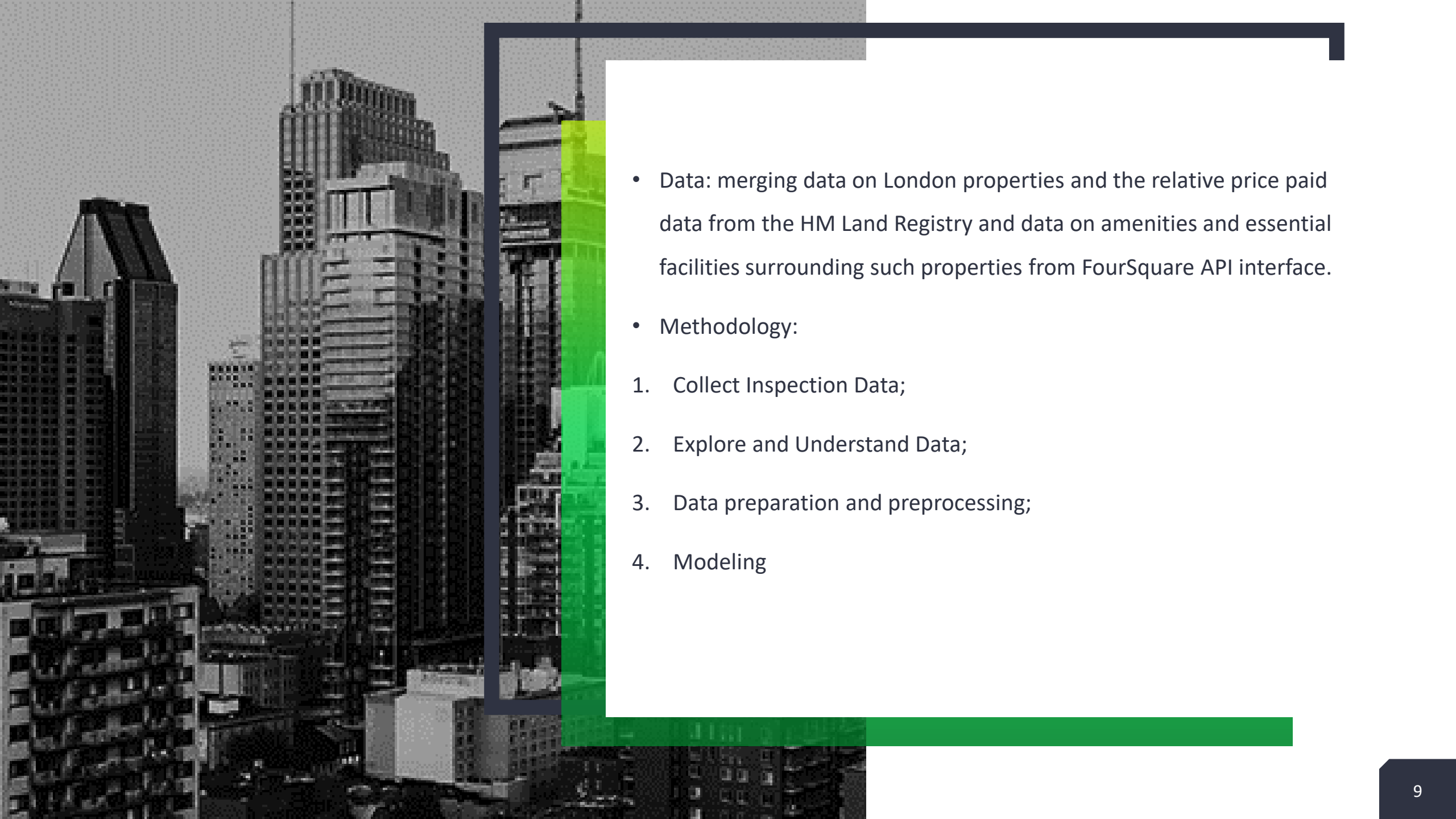
SOLUTION

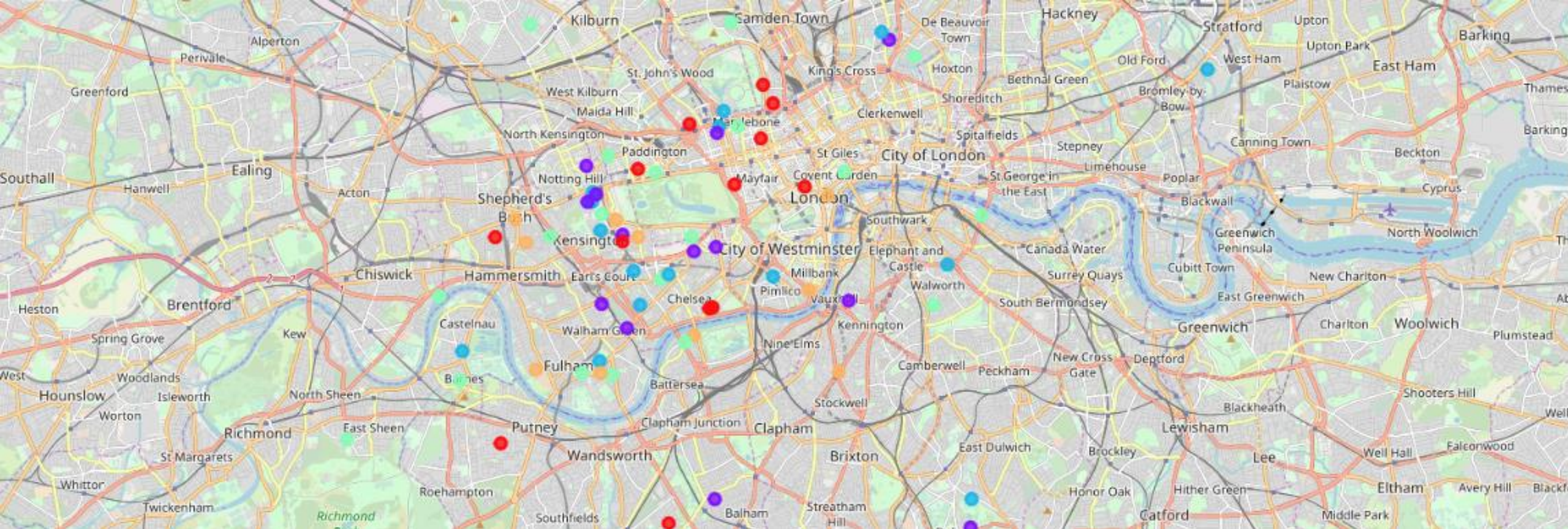


Clustering London neighborhoods in order to recommend venues and the current average price of real estate where homebuyers can make a real estate investment.



DATA AND METHODOLOGY

- 
- Data: merging data on London properties and the relative price paid data from the HM Land Registry and data on amenities and essential facilities surrounding such properties from FourSquare API interface.
 - Methodology:
 1. Collect Inspection Data;
 2. Explore and Understand Data;
 3. Data preparation and preprocessing;
 4. Modeling



K-Means clustering

The background is a grayscale aerial photograph of a city with numerous skyscrapers. Overlaid on this is a graphic consisting of a green rectangle on the left and a white rectangle on the right, separated by a thick black L-shaped line. The word 'OUTCOME' is centered in the white rectangle.

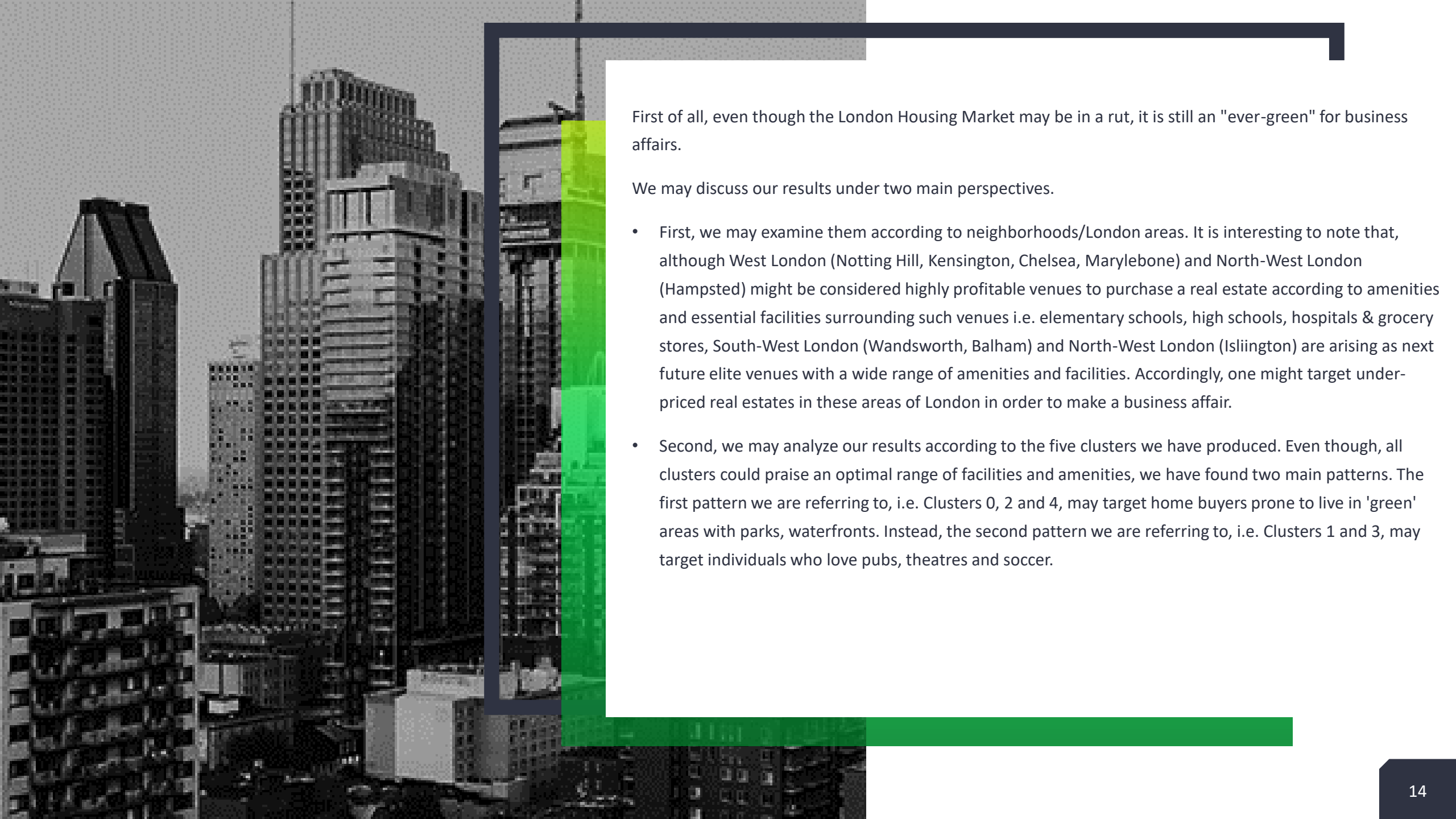
OUTCOME



- Examination of real estates according to neighborhoods/London areas
 1. West London (Notting Hill, Kensington, Chelsea, Marylebone) and North-West London (Hampsted) might be considered highly profitable venues to purchase a real estate;
 2. South-West London (Wandsworth, Balham) and North-West London (Islington) are arising as next future elite venues with a wide range of amenities and facilities.
- Examination of real estates by clusters
 1. Clusters 0, 2 and 4 may target home buyers prone to live in 'green' areas with parks, waterfronts;
 2. Clusters 1 and 3 may target individuals who love pubs, theatres and soccer.

The background is a grayscale aerial photograph of a dense urban area with numerous skyscrapers. Overlaid on this is a graphic consisting of a green rectangle with a black L-shaped border. Inside the corner of this border is a white rectangle containing the text.

RESULTS AND DISCUSSION




First of all, even though the London Housing Market may be in a rut, it is still an "ever-green" for business affairs.

We may discuss our results under two main perspectives.

- First, we may examine them according to neighborhoods/London areas. It is interesting to note that, although West London (Notting Hill, Kensington, Chelsea, Marylebone) and North-West London (Hampstead) might be considered highly profitable venues to purchase a real estate according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores, South-West London (Wandsworth, Balham) and North-West London (Islington) are arising as next future elite venues with a wide range of amenities and facilities. Accordingly, one might target under-priced real estates in these areas of London in order to make a business affair.
- Second, we may analyze our results according to the five clusters we have produced. Even though, all clusters could praise an optimal range of facilities and amenities, we have found two main patterns. The first pattern we are referring to, i.e. Clusters 0, 2 and 4, may target home buyers prone to live in 'green' areas with parks, waterfronts. Instead, the second pattern we are referring to, i.e. Clusters 1 and 3, may target individuals who love pubs, theatres and soccer.



CONCLUSION



To sum up, according to Bloomberg News, the London Housing Market is in a rut. It is now facing a number of different headwinds, including the prospect of higher taxes and a warning from the Bank of England that U.K. home values could fall as much as 30 percent in the event of a disorderly exit from the European Union. In this scenario, it is urgent to adopt machine learning tools in order to assist homebuyers clientele in London to make wise and effective decisions. As a result, the business problem we were posing was: how could we provide support to homebuyers clientele in to purchase a suitable real estate in London in this uncertain economic and financial scenario?

To solve this business problem, we clustered London neighborhoods in order to recommend venues and the current average price of real estate where homebuyers can make a real estate investment. We recommended profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores.

First, we gathered data on London properties and the relative price paid data were extracted from the HM Land Registry (<http://landregistry.data.gov.uk/>). Moreover, to explore and target recommended locations across different venues according to the presence of amenities and essential facilities, we accessed data through FourSquare API interface and arranged them as a data frame for visualization. By merging data on London properties and the relative price paid data from the HM Land Registry and data on amenities and essential facilities surrounding such properties from FourSquare API interface, we were able to recommend profitable real estate investments.

Second, The Methodology section comprised four stages: 1. Collect Inspection Data; 2. Explore and Understand Data; 3. Data preparation and preprocessing; 4. Modeling. In particular, in the modeling section, we used the k-means clustering technique as it is fast and efficient in terms of computational cost, is highly flexible to account for mutations in real estate market in London and is accurate.

Finally, we drew the conclusion that even though the London Housing Market may be in a rut, it is still an "ever-green" for business affairs. We discussed our results under two main perspectives. First, we examined them according to neighborhoods/London areas. although West London (Notting Hill, Kensington, Chelsea, Marylebone) and North-West London (Hampstead) might be considered highly profitable venues to purchase a real estate according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores, South-West London (Wandsworth, Balham) and North-West London (Islington) are arising as next future elite venues with a wide range of amenities and facilities. Accordingly, one might target under-priced real estates in these areas of London in order to make a business affair. Second, we analyzed our results according to the five clusters we produced. While Clusters 0, 2 and 4 may target home buyers prone to live in 'green' areas with parks, waterfronts, Clusters 1 and 3 may target individuals who love pubs, theatres and soccer.



THANK YOU