|  |  |
| --- | --- |
| NEW CINEMA LOCATION | |
|  | Film Reel Stock Photos, Pictures & Royalty-Free Images ... |
| BATTLE OF THE NEIGHBOURHOODS CAPSTONE DATA SCIENCE PROJECTAugust 2021Catherine Brown |  |

|  |
| --- |
| **Week 1 Task:**  *For this week, you will required to submit the following:*   * *A description of the problem and a discussion of the background. (15 marks)* * *A description of the data and how it will be used to solve the problem. (15 marks)*  DESCRIPTION OF THE PROBLEM & DISCUSSION OF THE BACKGROUND *Scenario:*  I have been tasked by a client with finding a suitable location in the Nottinghamshire area for a new cinema to be built or for an existing building to be converted. The Nottinghamshire area already has many cinema venues, so the location has to be carefully selected to ensure that it will be a successful business venture. My client already has a chain of independent cinemas in the North of England and is looking for a new location in the midlands, with Nottinghamshire being the preferred location.  *The scope of this project within the wider business analysis:*  Opening a new cinema requires significant capital investment. Deciding on a location, whilst an important part of the strategic decision, is only one of many factors to be explored. The local demographics of the area is an important factor to consider. Other analysis such as SWOT and 5 Market forces should also be a starting point for decision making. The barriers to entry for building a new cinema are high due to the upfront capital investment and the understanding of the business sector. Streaming providers such as Netflix and Amazon Prime should be considered as a threat to the viability of existing and new entrants to the cinema market. The high supply cost of showing new release Hollywood movies should be considered when deciding what type of cinema to build. The projected revenues and costs should then be computed to determine the actual return on investment to ascertain whether the project is potentially viable. Funding would then need to be sourced. The scope of this project is to determine only which NG (Nottinghamshire) postcode is the most suitable location for a new cinema to be built.  *Types of cinemas:*  The existing cinemas in Nottinghamshire are Multiplex cinemas such as Odeon and CineWorld, along with smaller chains such as the Savoy cinemas. There is also a popular independent cinema, Broadway, in the centre of Nottinghamshire that specialises in classics and alternative films. The type of cinema to be built would depend on the results of the business analysis. Cinemas such as the Savoy or Broadway are much smaller than the Multiplex cinemas and will perhaps show only a couple of the main release films at one time, offering older films that are cheaper from a supply perspective on the other screens. They offer a more intimate experience than the multiplex cinemas, with Broadway also having a bar and small restaurant. Creating more of an experience is likely to be ever more important when competing for customers with Netflix and Amazon Prime. DESCRIPTION OF THE DATA AND HOW IT WILL BE USED TO SOLVE THE PROBLEM *1. NG Postal code data:*  The Nottingham postal code data will be extracted from the webpage: <https://en.wikipedia.org/wiki/NG_postcode_area>’, using the function dfs=pd.read\_html(url) to read the data into a pandas dataframe. This will data will be used to select from when determining a suitable location for a new cinema. The extracted data will need to have coordinates found, so the postcode district and post town are combined into a full address column:     1. *Converting the NG Postal code data into coordinates for mapping:*   Using Geopy, the NG Postal code data will be used to find the approximate centre latitude and longitude coordinates of the postal codes. These coordinates will be added to the dataframe.     1. *Mapping the NG Postal codes*   The postal codes will then be mapped using Folium to provide a visual of the distribution of the postal codes.     1. *Obtaining existing Cinema venues around the Nottinghamshire area*   Foursquare will be used to obtain existing cinema venues around the Nottinghamshire area. This information will be stored in a dataframe, the data cleaned and the resulting location venues will be mapped onto the same map as the NG Postal codes. This will provide an initial visual of postal areas that do not already have a nearby cinema venue. It is better to identify a location that does not already have a cinema so that there is less competition for customers.     1. *Clustering the NG Postal codes using popular venues*   Foursquare will be used again to find the nearby 100 venues for each postal area. The data is then grouped by postal code area and sorted into the top 10 venues within each postal code area. A clustering method called KMeans will be applied to this data in order to cluster the postal code areas based on top local venues. To find the optimal k for the algorithm, the Elbow method is used. The information produced from this clustering exercise will provide an idea as to which postal code areas are similar types of Neighbourhoods based on nearby venues.       1. *Mapping the clusters on the same map and determining the best location*     With each area assigned to a particular cluster and imposed on the map, this will then provide an additional visual of which clusters tend to have cinemas and help identify if there are any postal codes that fall in the same cluster but that do not have a nearby cinema. Such a location could be a potentially good location for a new cinema.   1. *Limitations of the scope of this project*   It is important to note again that proximity to existing cinemas and similarity of neighbourhoods based on popular venues are only a very small part of the overall picture when deciding where to build a new cinema in Nottinghamshire and that the results of this project should be used in conjunction with a vast array of other supporting material as discussed in the wider business analysis section. |

**Week 2 Task (to be completed)**

*For the second week, the final deliverables of the project will be:*

* *A link to your Notebook on your Github repository, showing your code. (****15 marks****)*
* *A full report consisting of all of the following components (****15 marks****):*
* *Introduction where you discuss the business problem and who would be interested in this project.*
* *Data where you describe the data that will be used to solve the problem and the source of the data.*
* *Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.*
* *Results section where you discuss the results.*
* *Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.*
* *Conclusion section where you conclude the report.*
* *Your choice of a presentation or blogpost. (****10 marks****)*