Capstone Project - Presentation

The battle of Neighborhoods

Uy Du TRAN

1. Introduction

Background: Safety is always a top concern when living in a new area, new city. Especially this is a one of the biggest city in the world like London.

Problem: Nobody want to live in a dangerous neighborhoods with high crime's ratio. Here we define where are the safest borough in London base on the total crime's ratio. You like coffee? Or French restaurant nearby your home ? This project also will lead you to explore the venues in all the neighborhoods in London for helping you to find out where you are interested in the most.

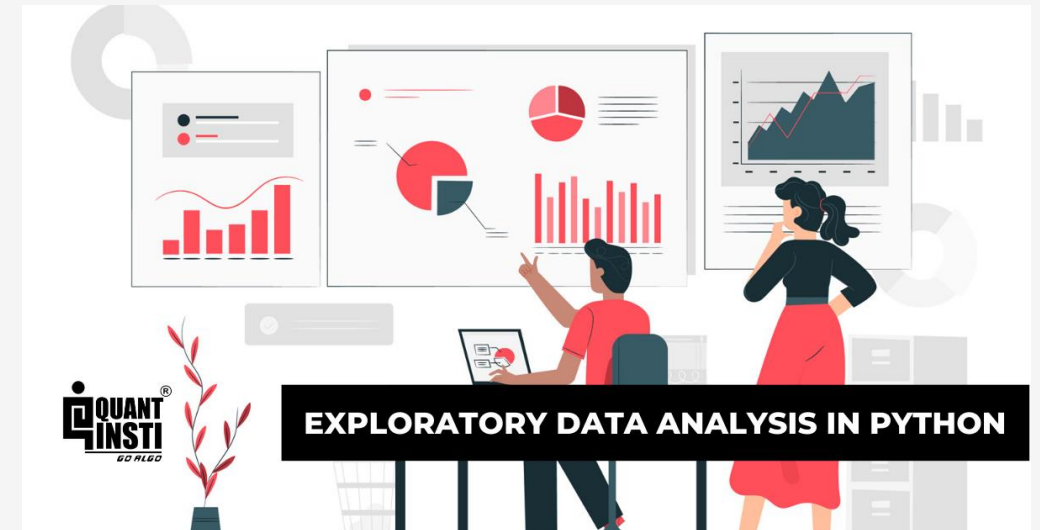Target: This project will serve two group of audience:

- Citizen

- Traveler

2. Analytic Approach

There are two approach in this project

1 *Exploratory data analysis(EDA)* to uncover hidden properties of data and provide useful insights to the reader

2 *Prescriptive analytics* to help people to explore all the neighborhoods to find out which is the borough the most suitable to live with *clustering* (KMeans)

EXPLORATORY DATA ANALYSIS IN PYTHON

3. Data requirement – Data acquisition

Data Acquisition: The data used in this project is a combination of data from:

1. The London crime data that shows the crime per borough in London. (Provide by: Kaggle)

2. The data from Wikipedia page that contains the list if London boroughs

3. The data acquired from the Foursquare API

	lsoa_code	borough	major_category	minor_category	value	year	month
0	E01001116	Croydon	Burglary	Burglary in Other Buildings	0	2016	11
1	E01001646	Greenwich	Violence Against the Person	Other violence	0	2016	11
2	E01000677	Bromley	Violence Against the Person	Other violence	0	2015	5
3	E01003774	Redbridge	Burglary	Burglary in Other Buildings	0	2016	3
4	E01004563	Wandsworth	Robbery	Personal Property	0	2008	6

```
<head>
  <meta charset="UTF-8"/>
  <title>
   List of London boroughs - Wikipedia
  </title>
  <script>
   document.documentElement.className = document.documentElement.className.replace( /(^|\s)client-nojs(\s|$)/, "$1client-js$2"
);
  </script>
  <script>
   (window.RLQ=window.RLQ||[]).push(function(){mw.config.set({"wgCanonicalNamespace":"","wgCanonicalSpecialPageName":false,"wgN
amespaceNumber":0,"wgPageName":"List_of_London_boroughs","wgTitle":"List of London boroughs","wgCurRevisionId":881899861,"wgRev
isionId":881899861,"wgArticleId":28092685,"wgIsArticle":true,"wgIsRedirect":false,"wgAction":"view","wgUserName":null,"wgUserGr
oups":["*"],"wgCategories":["Use dmy dates from August 2015","Use British English from August 2015","Lists of coordinates","Geo
graphic coordinate lists","Articles with Geo","London boroughs","Lists of places in London"],"wgBreakFrames":false,"wgPageConte
ntLanguage":"en","wgPageContentModel":"wikitext","wgSeparatorTransformTable":["",""],"wgDigitTransformTable":["",""],"wgDefault
DateFormat":"dmy","wgMonthNames":["","January","February","March","April","May","June","July","August","September","October","N
ovember","December"],"wgMonthNamesShort":["","Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec"],"wgReleva
ntPageName":"List_of_London_boroughs","wgRelevantArticleId":28092685,"wgRequestId":"XMbUCwpAAEUAACKj1V4AAABQ","wgCSPNonce":fals
e,"wgIsProbablyEditable":true,"wgRelevantPageIsProbablyEditable":true,"wgRestrictionEdit":[],"wgRestrictionMove":[],"wgFlaggedR
evsParams":{"tags":{}},"wgStableRevisionId":null,"wgMediaViewerOnClick":true,"wgMediaViewerEnabledByDefault":true,"wgPopupsRefe
rencePreviews":false,"wgPopupsConflictsWithNavPopupGadget":false,"wgVisualEditor":{"pageLanguageCode":"en","pageLanguageDir":"l
tr","pageVariantFallbacks":"en","usePageImages":true,"usePageDescriptions":true},"wgMFDisplayWikibaseDescriptions":{"search":tr
ue,"nearby":true,"watchlist":true,"tagline":false},"wgRelatedArticles":null,"wgRelatedArticlesUseCirrusSearch":true,"wgRelatedA
rticlesOnlyUseCirrusSearch":false,"wgMESchemaEditAttemptStepOversample":false,"wgPoweredByHHVM":true,"wgULSCurrentAutonym":"En
glish","wgNoticeProject":"wikipedia","wgCentralNoticeCookiesToDelete":[],"wgCentralNoticeCategoriesUsingLegacy":["Fundraisin
g","fundraising"],"wgWikibaseItemId":"Q6577004","wgCentralAuthMobileDomain":false,"wgEditSubmitButtonLabelPublish":true});mw.lo
ader.state({"ext.gadget.charinsert-styles":"ready","ext.globalCssJs.user.styles":"ready","ext.globalCssJs.site.styles":"read
y","site.styles":"ready","noscript":"ready","user.styles":"ready","ext.globalCssJs.user":"ready","ext.globalCssJs.site":"read
y","user":"ready","user.options":"ready","user.tokens":"loading","ext.cite.styles":"ready","mediawiki.legacy.shared":"ready","m
ediawiki.legacy.commonPrint":"ready","mediawiki.toc.styles":"ready","wikibase.client.init":"ready","ext.visualEditor.desktopArt
icleTarget.noscript":"ready","ext.uls.interlanguage":"ready","ext.wikimediaBadges":"ready","ext.3d.styles":"ready","mediawiki.s
kinning.interface":"ready","skins.vector.styles":"ready"});mw.loader.implement("user.tokens@0tffind",function($,jQuery,require,
module){/*@nomin*/mw.user.tokens.set({"editToken":"+\\","patrolToken":"+\\","watchToken":"+\\","csrfToken":"+\\"});
});RLPAGEMODULES=["ext.cite.ux-enhancements","site","mediawiki.page.startup","mediawiki.page.ready","jquery.tablesorter","media
wiki.toc","mediawiki.searchSuggest","ext.gadget.teahouse","ext.gadget.ReferenceTooltips","ext.gadget.watchlist-notice","ext.gad
get.DRN-wizard","ext.gadget.charinsert","ext.gadget.refToolbar","ext.gadget.extra-toolbar-buttons","ext.gadget.switcher","ext.c
entralauth.centralautologin","mmv.head","mmv.bootstrap.autostart","ext.popups","ext.visualEditor.desktopArticleTarget.init","ex
t.visualEditor.targetLoader","ext.eventLogging","ext.wikimediaEvents","ext.navigationTiming","ext.uls.eventlogger","ext.uls.ini
t","ext.uls.compactlinks","ext.uls.interface","ext.quicksurveys.init","ext.centralNotice.geoIP","ext.centralNotice.startUp","sk
ins.vector.js"];mw.loader.load(RLPAGEMODULES);});
  </script>
```

3. Data requirement – Data cleaning

The data acquired from these source are cleansed separately

1 The data is scraped from Wikipedia page using the Beautiful Soup library in order to extract the data in the tabular format as shown on the website.

2 The London crime data source, only crimes in the year 2016 is selected. We pivot the crime's data into different categories and get the total crimes per boroughs .

3 We merge these data by the borough's name to from the new dataframe. The purpose of the new data is for visualise the crime rates in each borough and identify with saftest borough during the year of 2016

4. Discussion

1 *The aim of this project is to help people to understand about the safest borough to live in London*

2 *The project will lead people to discover their favorite boroughs to live base on selecting nearby venues.*

Let's discover the project