

BIG DATA ANALYTICS AND VISUALISATION

DATA STORYTELLING PROJECT

BACKGROUND

Your team has been assigned a dataset to work on for the project.

DATASET PROVIDED

You are provided with a dataset “**BeijingHousePrices.xlsx**” on the housing prices in Beijing for the period from 2011 to 2017. The data is taken from Lianjia.com.

A download of the data (available in <https://www.kaggle.com/ruiqurm/lianjia>) can be retrieved from:

Canvas->Files->Projects.

Dataset Details

There are a total of 26 columns in the dataset. They are described here:

FIELD	DESCRIPTION
url	url of property (house)
id	Id of transaction
Lng	Longitude of property using the BD09 protocol
Lat	Latitude of property using the BD09 protocol
Cid	Community id
tradeTime	Date of transaction
DOM	Active days on market. More details in https://en.wikipedia.org/wiki/Days_on_market
followers	The number of people following/interested in the transaction
totalPrice	The total price of property (in Chinese Yuan and in ten thousands)
price	The average price (in Chinese Yuan) per square meters
square	The size of property (in square meters)
livingRoom	The number of bedroom
drawingRoom	The number of drawing room
kitchen	The number of kitchen
bathroom	The number of bathroom
floor	The number of floors in building. “High”, “Middle”, “Low”, “Top” are indicators of the position of the property unit in the building.
buildingType	1:tower, 2:bungalow, 3:combination of plate and tower, 4:plate
constructionTime	When was the property constructed
renovationCondition	1:other, 2:rough, 3:simplicity, 4:refined decoration
buildingStructure	1:unknown, 2:mixed, 3:brick and wood, 4: brick and concrete, 5:steel, 6:steel-concrete composite
ladderRatio	the proportion between number of residents on the same floor and number of ladders. It describes how many ladders a resident has on average.
elevator	0:does not have elevator, 1:has elevator
fiveYearsProperty	0:policy does not apply, 1: policy applies. Chinese government housing policy: if a person has owned a property for over five years before selling it AND if the property is the only real estate he owns in Beijing, then the Value Added Tax on this transaction can be waived.
subway	0:No subway nearby, 1:subway nearby
district	1:Dongcheng District, 2:Fengtai District, 3:Tongzhou District, 4:Daxing District, 5:Fangshan District, 6:Changping District, 7:Chaoyang District, 8:Haidian District, 9:Shijingshan District, 10:Xicheng District, 11:Pinggu District, 12:Mentougou District, 13:Shunyi District. 3 districts are not included in dataset: Yanqing district, Huairou district, and Miyun district
communityAverage	The number of people living in the community

WHAT ARE YOU REQUIRED TO DO?

Your team is the data analytics team of a large organization with capabilities in real estate investment, development, leases, and mall management. Your management has interest in venturing into Beijing properties. They provided your team with the data set as given above. Your team is required to analyze the dataset and draw insights into some business questions your team believes could be of interest to the management. Prepare a presentation of the insights you have gathered from your analysis and suitable recommendations to the management.

You may approach this presentation in this manner:

1. Consider one level-1 business question that need answering from the data analysis.
2. Add in any further sub-level questions for producing the appropriate insights for your story.
3. Conduct appropriate data cleaning.
4. Conduct exploratory analysis of the data with data visualisations featuring appropriate *representation* and *presentation* of data.
5. **Use Tableau, MS Excel, or Python packages to produce your data visualisations.**
6. Craft your presentation with the data visualisations and produce a **coherent data storytelling presentation** with recommendations for your story.

You may assume that the data is recent enough for your analysis and recommendations.

Note that data insights lead to the production of story, but they are not stories themselves. Your team may seek out other **relevant** data to supplement the currently available dataset in your presentation.

DELIVERABLES

1. A PowerPoint (or similar) storytelling presentation.
 - a. There is no restriction on the number of slides but note that your team has only 10 minutes for the presentation.

SUBMISSION

1. Zip the PowerPoint file and name it as “**Team_XX.zip**” (for example, file name for team 9 is “Team_09.zip” and for team 19 it is “Team_19.zip”). **PLEASE FOLLOW THIS NAMING CONVENTION ACCORDINGLY.**
2. Upload the zip file to “Projects” folder in Canvas->Assignments **by 17 July 2024 2359hr.**
3. **Each member of the team must submit a copy of the same team zip file to the canvas assignment folder.**

EVALUATION

1. Each team is given 10 minutes to present and 5 minutes Questions and Answers during the allocated 20 minutes Project Evaluation timeslot.
2. Your team will be evaluated on the data visualisation and storytelling presentation.
3. Each member in the team is required to present, even for a minute.
4. Evaluation criteria:
 - a. **Business Questions to Answer (15 marks):**
 - i. Define very clear, important, and interesting business questions that can be answered by visualizations.
 - b. **Knowledge/Understanding (15 marks):**
 - i. Demonstrate full knowledge of the subject of data visualization and dataset.
 - ii. Understanding of the dataset.
 - iii. Accurate use and perspective of the dataset.
 - c. **Story (12 marks):**
 - i. A convincing, coherent story crafted from data insights.
 - d. **Visualization (12 marks):**
 - i. Visualization diagrams are used professionally to give informative presentations.
 - e. **Layout (12 marks):**
 - i. **Colors:** Use of different and appropriate color of each diagram.
 - ii. **Neatness:** Very neat and clearly readable diagrams.
 - iii. **Title and accurate labels:** Clear title for the charts. Labels located correctly. No spelling errors.
 - f. **Creativity (12 marks):**
 - i. Synthesizing novel ideas to reach creative insights.
 - g. **Presentation style (12 marks):**
 - i. Professional presentation by presenters.
 - ii. Appropriate dressing
 - h. **Questions and Answers (10 marks):**
 - i. Ability to respond to questions posed.