**Introduction**

So far in this course, you have gone through modules on data visualization and storytelling that gave you knowledge on how to visualize the data and communicate your findings and insights from the data to technical and non-technical business stakeholders. In this assignment, you will put together everything you have learned in this course by visualizing a real data set on Airbnb and incorporating the best practices of dashboarding and storytelling in Tableau.  
   
Airbnb is an online American marketplace that allows property owners to list their properties on the platform and connect with travelers who are looking for a place to stay. Suppose you are working as a data analyst at Airbnb in Amsterdam. You are asked to analyze the landscape of Airbnb in the city and share some insights with your managers.  
   
You will create dashboards in Tableau as well as analyze the data in Python to discover interesting insights for the Airbnb listings in the city of Amsterdam. After that, you need to document these insights in an executive summary and share them with your managers. Here’s the overall breakup of the assignment deliverables and their allocated weights.

* Assignment Part I - Tableau Dashboards (45%)
* Assignment Part II - Python Analysis (45%)
* Executive Summary (10%)

Data

Internally, Airbnb collects data from its website to understand how they affect the residential real estate rental market. The data are freely available for public use [here](http://insideairbnb.com/get-the-data.html)

For this assignment, we need the listings data for the city of Amsterdam. We have already collected a subset of the raw data available on the aforementioned website.

You’ll be using this dataset for performing all the analyses and dashboards for this assignment.

Here's [the data dictionary](https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHlNyGInUvHg2BoUGoNRIGa6Szc4/edit#gid=982310896) for you to understand the different columns in the dataset. As mentioned earlier, the dataset that we have provided you is a subset of a much larger dataset. Hence you may find a lot more columns in the data dictionary as compared to the actual dataset provided above.

Disclaimer: This learning module may contain images, both still and moving, including photographs, advertisements, third-party trademarks, video clips etc. that may constitute copyrighted material procured from open-source, public domain, social media, or other print or digital publications, with the sole and bonafide intention of imparting education, disseminating information and illustrating an academic concept or an issue. We believe that this constitutes a 'fair dealing' exception under Section 52 of the Indian Copyright Act, 1957.

**Assignment Part I - Tableau Dashboards**

In the first part of the assignment, you’ll be using Tableau to analyze the listings data. Your overall objective is to prepare the following dashboards:

**Overview Dashboard I**

Create a dashboard with an overview of Airbnb for your city. This dashboard should give the reader an overall perspective of Airbnb. It should allow the user to choose a neighborhood and the dashboard should **at the least display the following aspects**:

* A Map Chart showing the listings in the chosen neighborhood
* List of summary metrics like number of listings, the total number of reviews, average rating, and average price
* Distribution of different types of rooms available in the neighborhood.

A screenshot of a map

AI-generated content may be incorrect.Here’s a sample dashboard that you can use as a reference to build your own.

**Note**: This is a template visualization and you are free to build your own version of the dashboard. Please make sure that all the relevant information as required in the problem statement is available in your dashboard.

**Overview Dashboard II**

Create another dashboard that dives deeper into the characteristics of the listings in Amsterdam. Here again, the user should be able to choose a neighbor and your dashboard should **at the least display the following aspects**:

* A chart showing the growth of Airbnb listings across the different years in the neighborhood.
* A chart displaying the**overview of top hosts** in the neighborhood.
* Pricing distribution across the different room types in the neighborhood.

Again, here’s a sample dashboard that you can use as a reference to build your own.

A screenshot of a graph

AI-generated content may be incorrect.

**Note:** This is a template visualization and you are free to build your own version of the dashboard. Please make sure that all the relevant information as required in the problem statement is available in your dashboard.

Charts and Dashboards guidelines

The charts should be informative to your analysis question and also demonstrate an understanding of the principles of data visualization covered in the modules. The guidelines for the charts are as follows:

* Each dashboard should have at least **2 distinct types of charts**, so each chart must contain a different type of visual that we covered in the modules.
* Each chart should be appropriately labeled to increase clarity.
* Color must be used consistently throughout the dashboard.
* Instructions and captions with the data source must be included in the dashboards.
* **IMPORTANT** - Your primary focus should be on ensuring that your dashboards contain the necessary charts that display the information mentioned in the problem statement above. However, you are encouraged to be creative with your dashboarding skills and improve the visualizations even further.

Each dashboard has different requirements, although there are some common themes. For each dashboard, we expect you to follow these guidelines at a minimum:

* Include informative and dynamic title(s) for the dashboard and/or chart and all appropriate chart elements such as axes.
* Reduce clutter to create an attractive and informative dashboard.
* Use color judiciously, including the background, gridlines, labels, bars/shapes, and so on.
* Implement dynamic elements so the reader can explore information about the neighborhoods.
* Include a textbox with introductions/instructions and a caption for the data source.

**Assignment Part II - Python Analysis**

In the next part of the assignment, you’ll be analyzing the data further using the tools and techniques you have learned in the EDA I and II modules.

The detailed steps of analysis and tasks have been mentioned in the stub file below. Please download the same and answer the questions pertaining to the different tasks.

Note: You have to use the same dataset as mentioned in the introduction page for peforming the tasks in this assignment.

**Assignment Part II - Stub File**

**Download**

Python Analysis Guidelines

* Go through the data dictionary thoroughly before starting the assignment. That will give you a good idea of what each column represents before you begin the analysis.

* Read each instruction carefully, identify the task to be performed, and only then write the required code. The assignment is meant to be straightforward. You do not need to perform additional analyses that are not requested explicitly. However, you are encouraged to perform a few additional analyses to get deeper into the insights

* Some of the tasks might require using functions you may not have used previously. In such cases, you should rely on the library documentation you referred to in the modules. Please understand that completing this assignment is a learning process, and research is part of it.

* Always run the cells of the Notebook sequentially, restart the kernel, and run all the cells to avoid runtime errors.

* Many of the questions will require you to view them from multiple angles. **You have been asked to observe any trends in the visualizations and provide insights for these trends.**In other words, there will be no fixed answers. You are expected to apply your problem-solving skills to come up with solutions and also document your work appropriately; both of these are part of the assignment grading.

**Executive Summary**

After performing the Tableau and the Python analysis, your final task is to summarize the insights from the dashboards and the Jupyter notebook and document them in an executive summary.

* The document should be structured and have a proper narrative regarding the Airbnb listings in the city of Amsterdam
* All the necessary visualizations from the Tableau dashboards and the Python Notebook should be present in the summary document.
* Set of insights/findings/recommendations for each dashboard, along with snapshots in pointers (max. 50 words each)
* Also, include your insights and observations from the analysis performed in Python
* A paragraph summarizing all the above-mentioned points at a top-level (max. 50 words)
* All the above should not be more than 500-600 words.

**Evaluation Rubric**

Below are the grading weights for each of the tasks.

|  |  |
| --- | --- |
| **Deliverable** | **Weightage** |
| Assignment Part I (Tableau Dashboards) | 45% |
| Assignment Part II (Python Analysis) | 45% |
| Executive Summary | 10% |

Detailed Rubrics

**Below are the grading weights for each of the dashboards (overall weightage - 45%)**

|  |  |  |
| --- | --- | --- |
| **Criterion** | **Meets Expectations** | **Does Not Meet Expectations** |
| Are your charts accurate according to the data? (40%) | * Uses the correct and appropriate variables * Aggregates the variables correctly | * Uses incorrect and/or inappropriate variables * Incorrectly aggregates variables (e.g., adds rather than counts) |
| Are your visualization choices appropriate for the analysis and data? (40%) | * Uses the correct chart given the data types and/or question * Uses the correct chart attributes (e.g., color palette) given the chart and data types | * Uses the incorrect chart given the data types and/or question * Improperly uses the chart attributes (e.g., color palette) given the chart and data types |
| Are your charts attractive, easy-to-read, well-labeled, and easy-to-understand? (5%) | * Chart labels and attributes are clear * Consistent use of visual elements (e.g., color, font) * Visual clutter such as grid lines and unnecessary text is removed * Strategic use of charts, ensuring they are large enough to read comfortably | * Chart labels and attributes are not clear * Inconsistent use of visual elements (e.g., color, font) * There is visual clutter such as grid lines and unnecessary text * Includes too many charts; charts are small and difficult to read |
| Are the title and focus of the dashboard clear? (5%) | * Reader can easily identify the purpose of the dashboard without knowing the assignment | * Reader cannot easily identify the purpose of the dashboard without knowing the assignment |
| How well-supported are your conclusions and recommendations? (5%) | * Conclusions and/or recommendations are relevant to the dashboard’s purpose * Conclusions and/or recommendations are naturally derived from the charts included in the dashboard | * Conclusions and/or recommendations are irrelevant to the dashboard’s purpose * Conclusions and/or recommendations are not shown in the charts included in the dashboard |
| Did you put in effort above and beyond the minimum? For example, did you create more sophisticated and/or interesting dashboards? (5%) | * Charts are sophisticated and/or visually interesting * Charts are complicated to create (e.g., uses filter actions, buttons, animations) * Dashboards are engaging and complex * Dashboards are thoroughly decluttered and coherent | * Charts are simple and/or visually boring * Charts are simple to create * Dashboards are not engaging or complex * Dashboards are cluttered and incoherent |

**Rubrics for the Python Notebook (overall weightage - 45%)**

| **Criteria** | **Meets Expectations** | **Does Not Meet Expectations** |
| --- | --- | --- |
| **Tasks #1–#3 (10% each)** | * Commands are syntactically correct. * Code output correctly answers the question. * Code is efficient; it uses built-in functions and libraries, when possible, rather than basic Python constructs (e.g., loops). | * Code has minor or major syntactical errors. * The functions used are only partially correct. * Code is inefficient, relying on basic Python constructs when built-in functions or libraries could have been used. * Code output does not directly or indirectly answer the question. * Results are unclear on how they address the tasks relevant to them. |
| **Tasks #4–#7 (15% each)** | * Commands are syntactically correct. * Code output correctly answers the question. * Code is efficient; it uses built-in functions and libraries, when possible, rather than basic Python constructs (e.g., loops). | * Code has minor or major syntactical errors. * The functions used are only partially correct. * Code is inefficient, relying on basic Python constructs when built-in functions or libraries could have been used. * Code output does not directly or indirectly answer the question. * Results are unclear on how they address the tasks relevant to them. |
| **Appropriate Markdown cells (5%)** | * Markdown cells are created to provide a high-level understanding of each analysis performed. * Markdown cells have no spelling errors and are written for the reader’s ease of understanding. * Markdown cells are well-formatted. | * Markdown cells are not used or are written poorly, and provide an ambiguous explanation of the analyses. * Markdown cells have spelling errors and/or the language is difficult to understand. * Markdown cells have not incorporated formatting elements to make reading easier. |
| **Appropriate code commenting and code style (5%)** | * Code comments cover all code written. * Code comments describe why code has been written. * Variable names are descriptive and unambiguous. * Code is structured for ease of reading. | * Code comments are either not used or only reiterate what the code is doing. * Variable names are vague. * Code is poorly structured, thus affecting readability. |

**Detailed Rubrics for the Executive Summary (10%)**

* The document should be structured and have a proper narrative regarding the Airbnb listings in the city of Amsterdam
* All the necessary visualizations from the Tableau dashboards and the Python Notebook should be present in the summary document.
* Set of insights/findings/recommendations for each dashboard, along with snapshots in pointers (max. 50 words each)
* Also, include your insights and observations from the analysis performed in Python
* A paragraph summarizing all the above-mentioned points at a top-level (max. 50 words)
* All the above should not be more than 500-600 words.

**Submission Guidelines**

Please read these guidelines carefully and ensure you are abiding by them at all times.

Do not make any changes to the original data set. Directly connect the data source to Tableau/Python and create the visualizations/perform the tasks. Do not modify the data set using Excel (or any other tool) and then use it in Tableau/Python. All data processing must be done in Tableau and Python only.

**Submission Guidelines**

**You have to make a submission of 3 files**

* **Tableau Dashboards:** You can either:
  + **Method 1** - Share the packaged workbook of your Tableau dashboards, or
  + **Method 2** **(Recommended)**- Publish the workbook to the Tableau public, and your presentation will be graded on the published Tableau workspace. To publish the workbook to the Tableau public, you will need to create the packaged Tableau workspace for submission. Go to Server > Tableau Public > Save to Tableau public.
* **Python Notebook:**
  + Submit the completed Python Notebook (with the code and observations/answers for each of the tasks
* **Executive Summary Document**: Submit the executive summary document as a PDF and it must include the following:
  + Dashboard link (in case you're sharing your Tableau dashboards through Tableau Public)
  + Brief summary of the dashboards about dimensions and measures used in each chart, calculations made, and filters applied (max. 100 words each)
  + Set of insights/findings/recommendations for each dashboard, along with snapshots in pointers (max. 50 words each)
  + Also include your insights and observations from the analysis performed in Python
  + A paragraph summarizing all the above mentioned points at a top-level (max. 50 words)
  + All the above should not be more than 500-600 words.
* Put all the above files in a folder, zip them and submit it in the link provided below