

**D09 - Data Visualization**

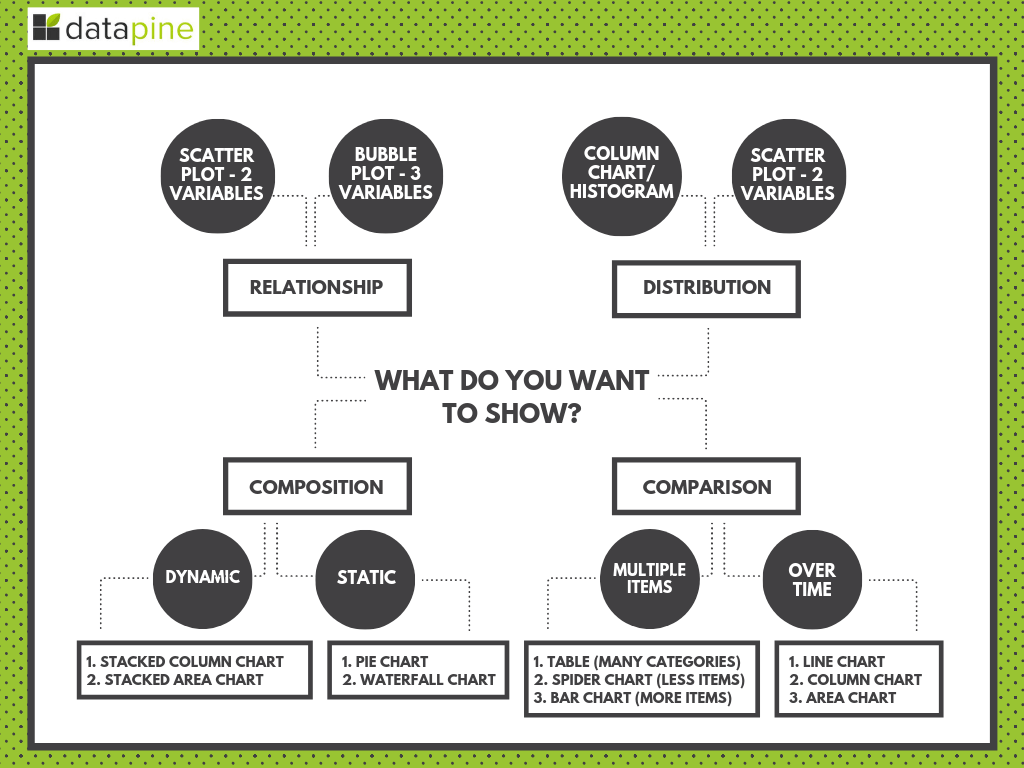
**Power BI**

As mentioned before on D01 and D04, our team used Power BI, a Microsoft’s cutting-edge platform for interactive data visualizations, analytics and reporting. Power BI offers a wide range of visualization types (line graphs, scatter plots, area diagrams, mapping etc.) and that’s why we tried to make use of the majority of these choices. Also, this platform fits perfectly for different business environments by providing the live dashboard option for engaging, real-time presentations.

As a company with high business standards and long-term perspective, we also deemed as necessary to install some extra visuals, such as WordCloud and Animated Bar Chart Race, which are available on Microsoft’s AppSource, by creating firstly a Power BI account. These data viz features gave a visually appealing tone to our project and lead to better decision-making. For example, by creating a WordCloud based on clients’ comments, we extracted useful insights for the properties and dived deeper into every line of the database in order to end up to more concrete conclusions for future business needs.

**Methodology**

For creating all the above-stated data visualizations and analytics, we adhered to the following methodology:



**Image 1**: [Data Visualization Methodology/ Source: <https://www.datapine.com/blog/how-to-choose-the-right-data-visualization-types/>]

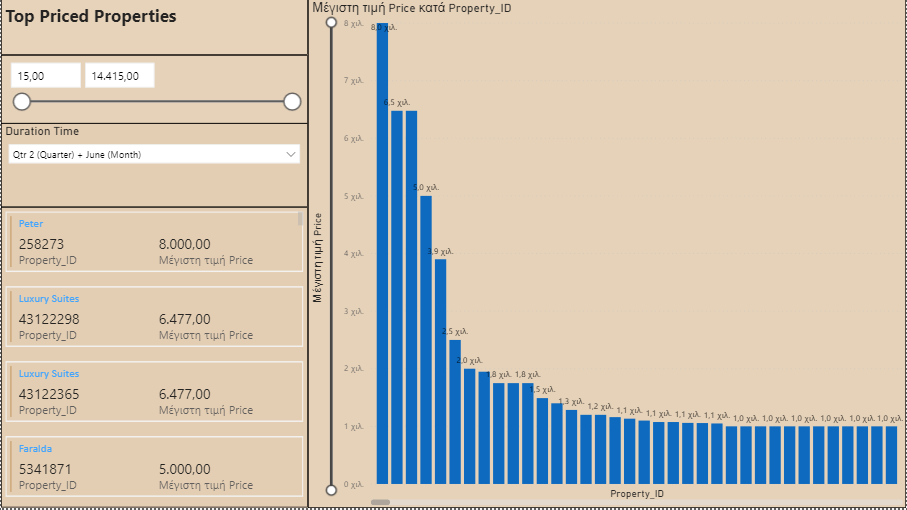
This diagram guided us on which should be our visualization type based on the need that we have to meet each time. It basically separates the data viz options into four main categories: relationship, distribution, composition and comparison. Each of these categories have also some subcategories which explain exactly why we should choose e.g. a scatter plot from bubble plot. In addition, there are some more criteria which actually shape the final dashboard, such as how dynamic it is, if it has multiple items (that is more than three) or how important is the “time” factor on the dataset examined.

**Data Visualization Types (by business need)**

Most of the asked business needs were finally answered - and then visualized - in various ways since we thought that there were not univocal answers or singular approaches. So, we aimed to examine these questions in a more critical and spherical way so as to be fully prepared for the client’s requirements.

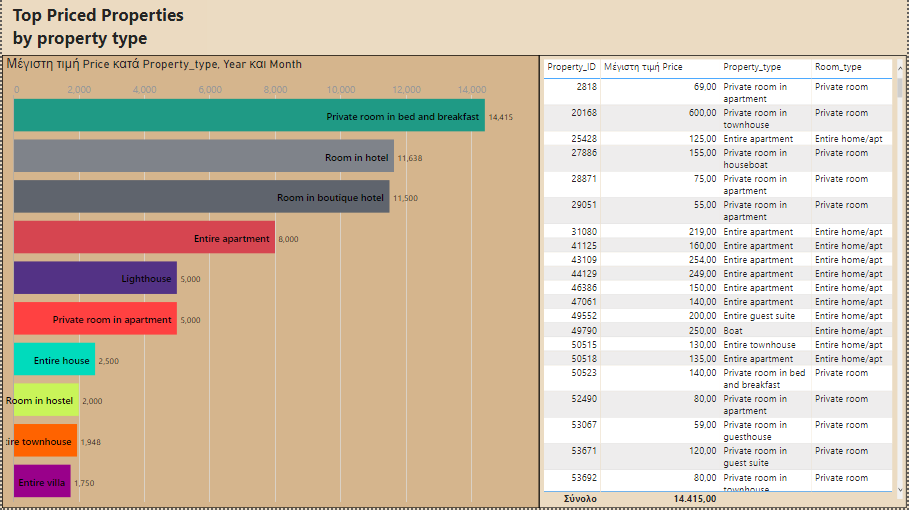
1. Which are the company’s top-priced properties?

The two chosen visualizations (stacked column chart and leaderboard) present the maximum price of each property based on its id. This dashboard is also framed by two helpful slicers: the first one adjusts the property’s price and the second regulates the “time” factor, both monthly and quarterly.



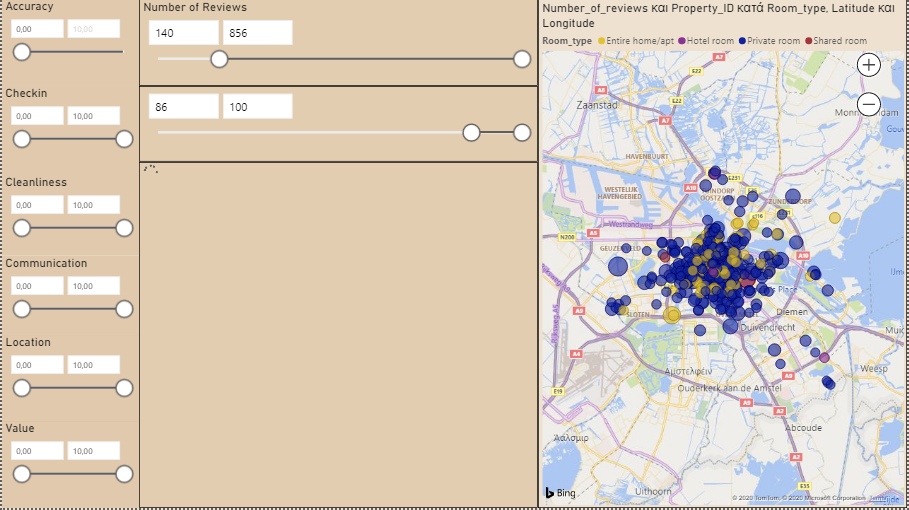
**Image 2**: Max Price & Property\_ID (1st Dashboard Approach)

This animated bar chart race (a) shows interactively how the maximum price of each property type (private room in bed and breakfast, room in hotel, room in boutique hotel etc.) changes, that is increases or decreases, based on the year and month. The right table (b) is a general ranking of the properties based on their maximum price and id, which in the same time presents the type of the room and the property itself.



**Image 3**: a-> Max Price, Property\_ID, Year & Month , b-> Property\_ID, Max Price, Property\_Type & Room\_type (2nd Dashboard Approach)

1. Which are the company’s top-rated properties?



Focusing on the above dashboard, we see a field map visualization which is actually created by the number of reviews, the property\_ID and its room\_type. On the left side there is a wide range of slicers for accuracy, checkin, cleanliness etc., which affect the final outcomes.

Also, centrally on the dashboard we want to make clear who are the property owners (based on their property\_id and host\_name) and how much high they score on the number\_of\_reviews and review\_scores\_rating. Of course, these two rankings can also be controlled from the relevant slicers. All these metrics are extremely necessary for the final user because he is capable of adjusting them and finally get meaningful answers on his business questions.

1. Which hosts are the busiest?

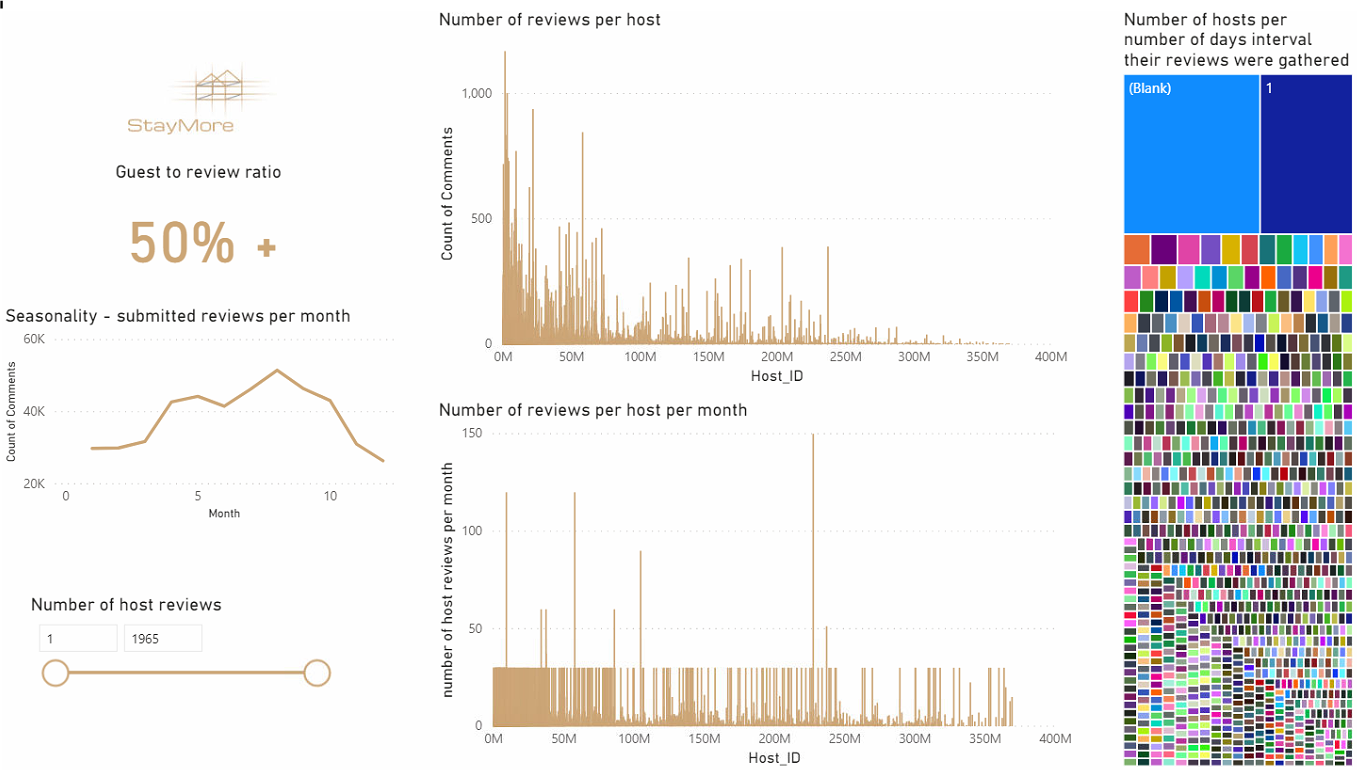
The following dashboard is a combination of different metrics which in fact depict how the number of reviews defines the busiest hosts.

On the left side, firstly it is placed a measure with the guest to review ratio (+50%). This is our assumption of how many property guests actually leave a review (50% or more is our assumption). Since we have no data for actual bookings happened, we assume that the reviews submitted for a property are a good measure of how busy a property is. By summing the reviews of all the properties belonging to a host, we can get a measure of how busy this host is.

Moreover, we discovered that there is a seasonality on reviews (based on submitted reviews per month) and hence on bookings as well, which would also translate to busiest hosts during the summer period where we see the peak.

We have also included two column charts, one displaying the number of total reviews received per host and another displaying the number of total reviews per host per month. The latter has been calculated by finding for each host, the minimum and maximum date of the reviews the host has received, deducting the min from the max and adding one, and then dividing this amount by the days a month has, in order to get a measure of the reviews per month each host receives on average.

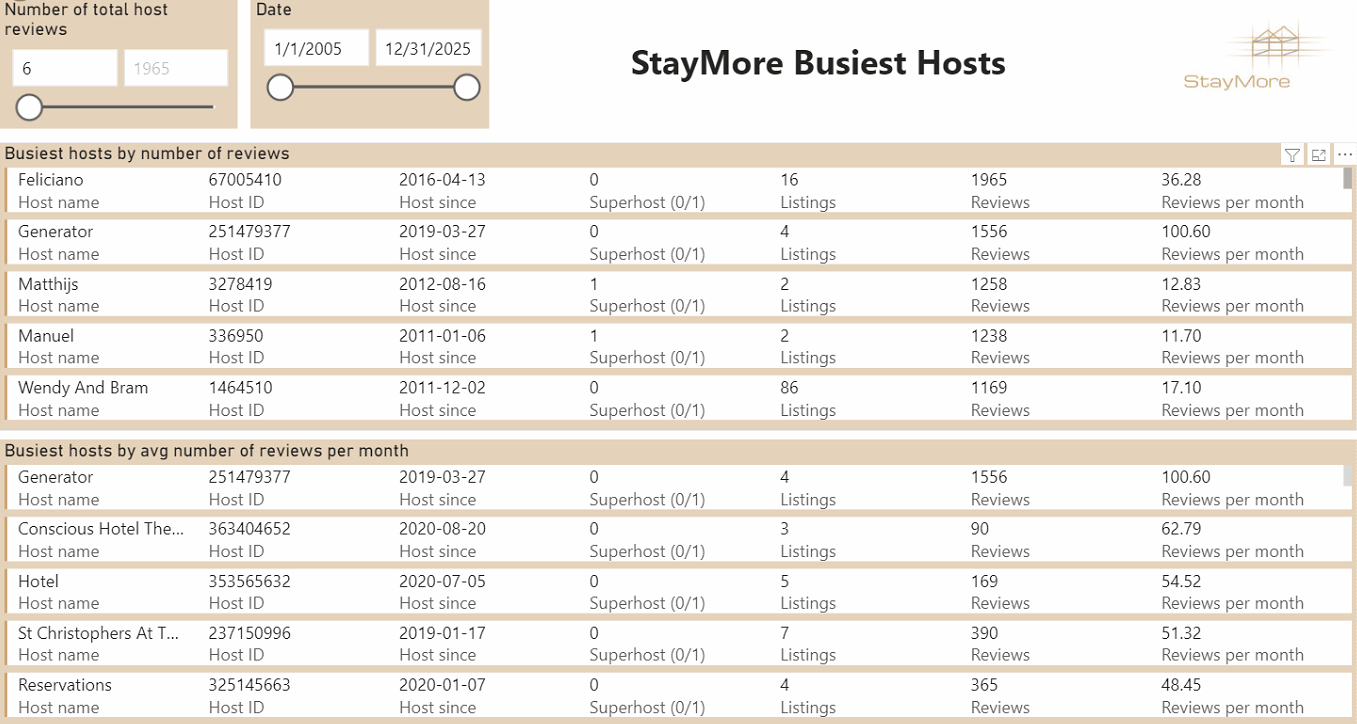
On the treemap, we have grouped the hosts by the time interval in which they have received reviews. We can see that there is a large group of hosts that haven’t received any reviews and the second largest group is of those hosts that have received reviews only in one day. The most popular category within this group is host that have only received 1 review (and naturally this is published in a single day). That has resulted in them having a reviews per month ratio of 30 which distorts our date. This is why we selected to use a slicer for the number of host reviews to filter out those hosts. When we set for instance the slicer to include hosts with 6 or more reviews only, the results improve significantly and we can get a more clear picture about the host reviews either by total number or the total number of reviews per month.



**Image 4**: Busiest Host (1st Dashboard Approach)

We have also created a second dashboard where the busiests hosts are presented in two leaderboards based on the two different metrics (total number of reviews or total number of reviews per month).

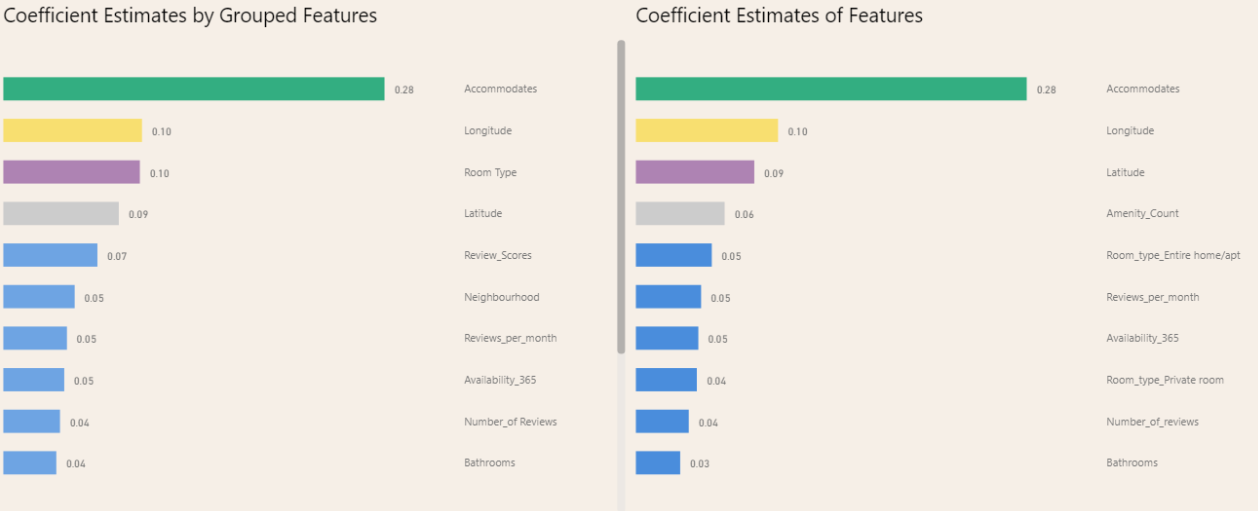
But, probably the most “friendly” tools on this dashboard are the two slicers because the user has the ability to adjust either the date in which they are searching for the busiests host or selecting an interval for the number of total host reviews.



**Image 5**: Busiest Hosts (2nd Dashboard Approach)

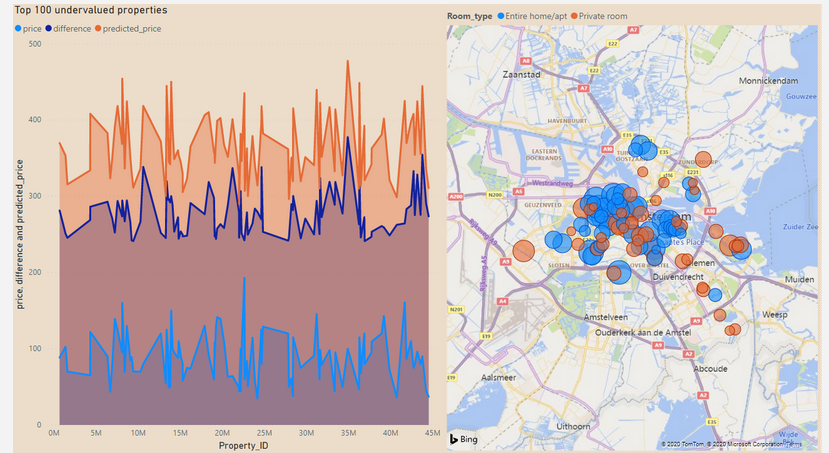
The first three business needs were answered by importing the data warehouse into Power BI, while the following four came as results after building a machine learning model:

1. Which attributes are the most critical for the price of a listing?



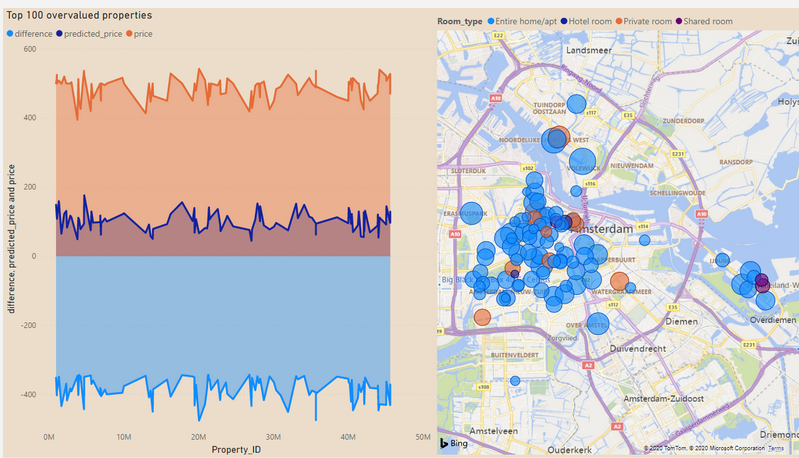
**Image 6**:

1. Which listings are the most undervalued?



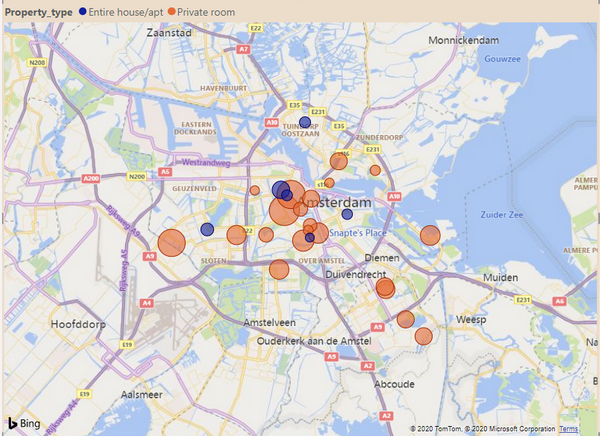
**Image 7**:

1. Which listings are the most overvalued?



**Image 8**:

1. Identify candidates based on their value and ratings whose hosts could be notified for increasing their price.



**Image 9**:

**Color Palette**

Visualization is all about aesthetics and communicating the right message to the viewer/business manager etc. So, the color palette that we chose for our visualization purposes was mainly professional looking and plenty of “business” colors (greys, blues). Of course, some extra vivid brushstrokes and gold background hues (e.g. #CCA677) were also used for pointing out the most crucial findings on our reporting.

Here are some of the main hues that we chose so as to make our visualizations more cohesive and good-looking:

* #003B46 (Deep Aqua)
* #07575B (Ocean)
* #66A5AD (Wave)
* #C4DFE6 (Seafoam)
* #1995AD (Glacier Blue)
* #A1D6E2 (Ice)
* #BCBABE (Warm Gray)
* #F1F1F2 (Overcast)