Hüseyin Soykök – Group 24

Visualisation Project - Domain Situation & Task Abstraction

Dataset: AB_NYC_2019 (New York City Airbnb Open Data).

Overarching Theme: Analysing the New York City (NYC) Airbnb market from the perspectives of its primary stakeholders: the traveller, the host, the regulator, the competitor and the analyst.

User Story 1: The Value-Seeking Traveler

Persona: Hüseyin, a prospective tourist planning a 5-day trip to NYC. He is budget conscious but also values popularity and safety (as indicated by reviews). He is unfamiliar with NYC geography and wants to find the best "price-performance"
accommodation.

Story: "Hüseyin is planning his first trip to New York and is overwhelmed by the accommodation costs. His budget is firm, between $100 and $150 per night. He only wants to stay in an 'Entire home/apt' and prefers listings that are popular and vetted, so he sets a filter for over 100 reviews. He needs to know: Where are these specific 'value-for-money' listings geographically clustered? Are there affordable neighbourhoods in Brooklyn or Queens that are safe (popular) and still close to Manhattan?"

Task Abstraction

High level: Discover (find new, unknown patterns) / Explore (find value hotspots).

Mid-level: Locate (find items on a map) / Identify (pick out targets).

Low level: Filter (by price < 150, number_of_reviews > 100, room_type == 'Entire home/apt') / Correlate (the filtered items with their latitude and longitude).

Target: Distribution (the geographic clustering) / Outliers (in a positive sense: the "value gems" he is seeking).

Final task: Hüseyin aims to Discover value hotspots by Filtering listings based on his criteria (price, reviews, room type) and then Locating the geographical Distribution of these specific targets on a map.

User Story 2: The New Host

Persona: David, a local resident who has saved some money and is considering buying a property to rent out on Airbnb. His goal is to maximize his potential return on investment (ROI).
Story: "David wants to invest in a property to list on Airbnb. He needs to decide which type of property in which area will be most profitable. He doesn't want to see individual listings, but rather the typical (median) price. He needs to know: What is the median price for an 'Entire home/apt' in Manhattan compared to an 'Entire home/apt' in Brooklyn? And how does that compare to a 'Private room' in Manhattan? He also wants to see the typical availability_365, as high availability might imply lower demand (occupancy)."

Task Abstraction

High level: Discover (find the most profitable segment) / Explore. Mid-level: Compare (compare market segments) / Summarize (find the typical price, not individual points).

Low level: Filter (by neighbourhood_group and room_type) / Aggregate (to calculate the median price and median availability_365).

Target: Distribution (the price distribution for each category) / Summary (the median, min, max for each segment).

Final task: David needs to Compare the summary Distribution (e.g., median price) across different categories (neighbourhood_group , room_type) to Discover the most profitable market segment to invest in.

User Story 3: The City Regulator

Persona: Dr. Chen, an urban planner working for the NYC city council. She is researching the impact of "commercial" or "ghost hotel" listings on the long-term housing market.

Story: "Dr. Chen is tasked with identifying listings that are likely not primary residences but full-time commercial operations. Her proxy for this is any listing that is available year-round (availability_365 > 300 days) OR is run by a 'super-host' who manages many properties (calculated_host_listings_count > 5). She needs to know: Where are these 'commercial-like' listings geographically concentrated? Are certain neighbourhoods (like 'Williamsburg' or 'Midtown') being disproportionately affected?"

Task Abstraction

High level: Discover (find patterns of commercial activity).

Mid-level: Identify (find the commercial listings) / Locate (find their concentration).

Low level: Filter (using a complex query: availability_365 > 300 OR calculated_host_listings_count > 5) / Correlate (these listings with their latitude / longitude).

Target: Distribution (the spatial clustering/concentration of these "commercial" listings).

Final task: Dr. Chen wants to Identify listings that meet a “commercial” criterion by Filtering the data and then analyse their spatial Distribution to Discover potential "hotspot" zones of high commercial activity.

User Story 4: The Hotel Competitor

Persona: Maria, the manager of a large chain hotel in Midtown, Manhattan. She is conducting a competitive analysis to understand the local Airbnb market.

Story: "Maria manages a hotel in 'Midtown'. She needs to understand her direct competition. She is not interested in listings in Brooklyn or Queens. She needs to know: Within my neighbourhood ('Midtown'), what is the range (distribution) of price for 'Private room' and 'Entire home/apt' listings? How does their number_of_reviews (popularity) compare to the average NYC listing? This will help her team position their own hotel rooms' pricing for the upcoming season."

Task Abstraction

High level: Explore (the local competitive landscape).

Mid-level: Compare (her neighbourhood vs. others) / Summarize (get the price range).

Low level: Filter (by neighbourhood == 'Midtown') / Aggregate (to get the price distribution and average number_of_reviews).

Target: Distribution (the spread of prices and reviews within her specific area).

Final task: Maria needs to Explore her local competition by Filtering the data to her specific 'Midtown' neighbourhood and then Summarize the Distribution of prices and review counts to Compare her hotel's positioning against nearby Airbnb listings.

User Story 5: The Data Journalist

Persona: Michael, a data journalist for a local NYC blog. He is writing a feature story about the "professionalization" of Airbnb in the city.

Story: "Michael's thesis is that Airbnb is no longer about 'sharing your home' but about professional 'mega-hosts' running a business. He needs data to support this. He needs to know: What is the distribution of listings per host? (i.e., How many hosts have just 1 listing vs. 10+ listings?). He also wants to see if the market's activity grew over time, using last_review dates to see when review activity peaked before this 2019 snapshot."

Task Abstraction

High level: Discover (market consolidation trends) / Analyse.

Mid-level: Compare (small hosts vs. mega-hosts) / Identify (temporal trends).

Low level: Aggregate (count hosts by calculated_host_listings_count) / Aggregate (count reviews by month/year from last_review).

Target: Distribution (of listings per host) / Trend (temporal activity from last_review dates).

Final task: Michael aims to Analyse market consolidation by Aggregating and Comparing the Distribution of listings per host, and to Identify temporal Trends by aggregating review dates.

Project Task Summary Table

Persona	Role	Core Question / Goal	Key Data Attributes	Primary Task (High-Level)
Hüseyin	The Value-Seeking Traveler	To find geographically clustered "value" listings (low price + high number_of_reviews).	price, number_of_reviews, room_type, latitude, longitude	Discover / Locate
David	The New Host (Investor)	To determine the most profitable market segment by comparing *typical* (median) price by region and room_type.	price (median), neighbourhood_group, room_type, availability_365	Discover / Compare
Dr. Chen	The City Regulator	To identify the geographical *concentration* (hotspots) of "commercial" listings (high availability or high host_listings_count).	availability_365, calculated_host_listings_count, latitude, longitude	Discover / Identify
Maria	The Hotel Competitor	To analyze the price *range* and *popularity* of direct competitors within her specific neighbourhood ('Midtown').	price (distribution), number_of_reviews, neighbourhood	Explore / Compare
Michael	The Data Journalist	To analyze the rise of "mega-hosts" (distribution of host_listings_count) and the market's *temporal* growth (via last_review dates).	calculated_host_listings_count, last_review	Analyze / Identify