

אוניברסיטת בן-גוריון בנגב  
Ben-Gurion University of the Negev

# פרויקט קורס ויזואליזציה

מלונות בהודו



מרצה : פיטר בק  
מגישה : הדר לקריץ  
ת.ז : 203954300  
תאריך : 01.03.2018

# תוכן עניינים

## תוכן

1.....	תוכן עניינים
2.....	What
2.....	הקדמה ותיאור התחום
2.....	תיאור הנתונים
4.....	ניקיון ועיבוד הנתונים
6.....	Why
6.....	תיאור המשתמש
6.....	User task
8.....	How
8.....	תהליך בניית הויזואליזציה
13.....	visual encoding
14.....	תוצאות הויזואליזציה
16.....	אינטראקטיביות הויזואליזציה
20.....	Value of visualization
20.....	יתרונות:
20.....	חסרונות:
21.....	הצעות לשיפור ושדרוג הויזואליזציה:
21.....	ערך הויזואליזציה לפי המדדים שנלמדו בכיתה:
24.....	דיון כללי

# What

## הקדמה ותיאור התחום

הודו הינה מדינה בדרום אסיה - השביעית בגודלה בעולם מבחינת שטח והשנייה בגודלה בעולם מבחינת אוכלוסייה. היא מונה למעלה ממיליארד נפשות. הודו נחשבת למדינה הדמוקרטית הגדולה ביותר בעולם.

הודו מהווה יעד טיול פופולארי בקרב מטיילים מכל העולם ובעיקר ישראלים. עשרות אלפי ישראלים טסים בכל שנה להודו - מטיילים צעירים אחרי צבא המחפשים מקום שונה, זול ומיוחד, צעירים בשנות ה-30 לחייהם המחפשים שקט רוחני, יוגה ומדיטציה וכן קבוצות של מטיילים מבוגרים יותר אשר לא עמדו בפיתוי והחליטו לבקר במקום שעליו כולם מדברים.

המפגש עם הודו הוא חוויה לטוב ואם זה לרע, במדינה הכל כך ייחודית, מלוכלכת, קסומה, מטונפת, מלאה רוחניות ועוני. הודו היא ללא ספק אחת המדינות המעניינות והמרתקות על הגלובוס, היא מעין מקרוקוסמוס של המזרח הרחוק ולמעשה היא השער למזרח עבור מטיילים רבים. היא מאגדת בתוכה שילוב וקסם של תרבויות שונות, עושר לשוני, נופים קסומים, המון רוחניות, וגם המון לכלוך, עוני קיצוני וחיים פשוטים וקשים.

החוויה בהודו מתבססת רבות על מקומות הלינה המגוונים בה. לכל אדם חשוב היכן יחזור לישון בלילה וזה יכול להוות את ההבדל בין מטייל שהתאהב בהודו למטייל ששונא אותה. הפערים בין העושר והעוני בהודו עצומים ועל כן האופציות ללינה בהודו בלתי מוגבלות. עקב כך, מטרת הפרויקט שלי הינה לפתח כלי ויזואליזציה אשר יצליח באופן קל ונוח לעזור למטייל להתאים לעצמו מקום לינה ראוי לפי התנאים בהם הוא חפץ, לפי מה שחשוב לו, הרכב הטיול שלו וסדר העדיפויות של המטייל. בנוסף, אני מייעדת את הויזואליזציה גם לאנשי עסקים שמעוניינים להשקיע במלון, לקנות או לבנות מלון בהודו. גם מבחינתם, הויזואליזציה תסדר ותציג את הנתונים בצורה מביאה תועלת.

## תיאור הנתונים

האתר "Kaggle" אתר המפרסם בסיסי נתונים חנים לצורכי ניתוח ולימוד, פרסם בסיס נתונים של מידע שנלקח מפורטל הודי בשם make my trip ומכיל אינפורמציה על מגוון מלונות בהודו. בסיס הנתונים המקורי שבאתר המלונות כולל 615,000 מלונות, אך בסיס הנתונים הניתן להורדה כולל 20,093 רשומות. בסיס הנתונים כלל טבלה יחידה (makeMyTrip) ובה השדות המפורטים מטה.

אנתח את כל המשתנים בטבלאות בהתאם למה שנלמד בכיתה-  
עבור כל משתנה בטבלה אתאר את המשמעות שלו ואפיין את הסוג שלו.

Attribute	Description	Data/Attribute Type
Area	Hotel's area in India	Categorical
City	Hotel's city in India	Categorical
Country	Hotel's country (India)	Categorical
highlight_value	What the hotel offers	Categorical
hotel_overview	Travel time to the airport, nearby places of entertainment, etc.	Categorical

hotel_star_rating	number of stars for the hotel	<b>Ordered Quantitative Discrete</b>	– –
image_urls	url for hotel's pictures	<b>Categorical</b>	
in_your_room	The contents of the rooms in the hotel	<b>Categorical</b>	
latitude	geographic coordinate that specifies the north–south position of a point on the Earth's surface	<b>Spatial</b>	
Longitude	geographic coordinate that specifies the east–west position of a point on the Earth's surface	<b>Spatial</b>	
mmt_holidayiq_review_count	The amount of visitors reviewed the hotel in holidayiq website	<b>Ordered Quantitative Discrete</b>	– –
mmt_location_rating	The location rating of the hotel (from 0 to 5)	<b>Ordered, quantitative, sequential</b>	
mmt_review_count	The amount of visitors reviewed the hotel (mmt_review_rating)	<b>Ordered Quantitative Discrete</b>	– –
mmt_review_rating	The amount of visitors reviewed the hotel in each of the categories: Excellent, Very Good, Average, Poor, Terrible	<b>Ordered Quantitative Discrete</b>	– –
mmt_review_score	Weighted average of mmt_review_rating	<b>Ordered, quantitative, sequential</b>	
mmt_traveller_type_review_count	The amount of visitors reviewed the hotel as suitable for each of the categories: Family, Couple, Solo, Friends, Business	<b>Ordered Quantitative Discrete</b>	– –
pageurl	The url of the hotel's page in "make my trip" website	<b>Categorical</b>	

property_address	The hotel's address in india	<b>Categorical</b>
property_name	The hotel's name	<b>Categorical</b>
property_type	Hotel, Homestay etc.	<b>Categorical</b>
room_types	Deluxe, Standard Non Ac, Suite etc.	<b>Categorical</b>
sitename	The name of the website (makeMyTrip)	<b>Categorical</b>
state	Name of the state where the hotel is located	<b>Categorical</b>
traveller_rating	Travelers rating for Location, Hospitality, Facilities, Cleanliness, Value for Money, Food (each from 0 to 5)	<b>Ordered, quantitative, sequential</b>

### ניקיון ועיבוד הנתונים

בוצעה עבודה רבה על הנתונים טרם תחילת העבודה על היוזאליזציה על מנת להצליח להשתמש בבסיס הנתונים שסיפק את המידע בצורה לא נוחה שרובה הורכבה מטקסט.

- תכונות שאינן רלוונטיות כלל ליוזאליזציה נמחקו מבעוד מועד – sitename (זהה לכל הרשומות), image\_urls, highlight\_value, hotel\_overview, in\_your\_room. השינוי נעשה ישירות על הנתונים באמצעות תכנת excel.
- רשומות חסרות נמחקו באמצעות שאילתת sql בשימוש בתכנת R.
- שדה traveler\_rating המורכב מדירוגים של 6 תכונות שונות (Location, Hospitality, Facilities, Cleanliness, Value for Money, Food) הופרד בעזרת תוכנת R ל-6 שדות נפרדים המכילים את הניקוד של כל תכונה כערך נומרי רציף.
- כל השדות המספריים שנקראו מהאקסל ל R כשדות פקטוריאליים עברו הסבה לשדות נומריים בדרכים שונות בשימוש ב R.
- שדה mmt\_traveller\_type\_review\_count הורכב מדירוגים של 5 תכונות שונות (Family, Couple, Solo, Friends, Business) הופרד בעזרת תוכנת excel (ובעזרת פיטר) ל-5 שדות נפרדים המכילים את מספר המדרגים של כל "סוג מלון" כערך נומרי נפרד.
- לאחר חשיבה על המשתמש ומה ירצה לראות בחיפוש, הוחלט כי הערך עצמו של השדות המסווגים מלון ל"סוג מטייל" (משפחה, חברים, זוג) הנותן מידע על כמות המדרגים לכל סוג לא רלוונטי, וכי הרלוונטיות הינה רק באם הערך של אחד מה"סוגים" הוא 0. זאת מכיוון שלא נרצה להסתיר מפני משתמש שמחפש מלון לזוג מלון מסוים רק מכיוון שיותר אנשים הצביעו

כי המלון מתאים למשפחות כאשר הוא עדיין מתאים גם לזוגות. רוב המלונות מכילים חדרים המתאימים למספר "סוגי מטיילים" ולא רק אחד. עקב כך בחרתי לעבור על כל ערכי 5 השדות המפורטים בנקודה מעל ולהחליף את הערכים ל-1 (במידה והמלון דורג לפחות פעם אחת כמתאים לסוג המטייל שבתכונה) או 0 (במידה ולא דורג מעולם כמתאים לסוג מטייל זה). כעת 5 התכונות הינן בינאריות ובהן אשתמש.

- כל שדות ה-character שנקראו מהאקסל ל R כשדות פקטוריאליים עברו הסבה לשדות char בשימוש ב R.

# Why

## תיאור המשתמש

יצרתי את הויזואליזציה כך שתוכל לשרת שני סוגי משתמשים:

1. עקב כך שהמטיילים הישראלים בהודו מגוונים כל-כך, בטווחי גיל שונים ומטיילים למטרות שונות, המשתמש הראשון בויזואליזציה זו הינו מטייל המתכנן טיול בהודו (באזור מסויים או במספר אזורים) ומעוניין לבחור לעצמו מקום לינה שתואם את צרכיו וסדר העדיפויות שלו. כדי לענות על מציאת המלון המתאים בצורה הטובה ביותר יצרתי מפה שתראה את המלונות המתאימים ביותר לצרכי המשתמש באמצעות כמה שיותר מיונים וסינונים כדי לתת מענה כמה שיותר מדויק. המפה תראה למשתמש את כל החלופות העומדות בפניו בהתאם לסינונים שבחר והוא יוכל להתמקד באזורים המעניינים אותו תוך כדי בחינת סביבת המלון (כבישים, נופים, שדות תעופה, הרים, נחלים ועוד) ותוך ראייה רוחבית של כל האופציות המתאימות ויכולת השוואה ביניהן.
2. משתמש נוסף שיכול להפיק תועלת מויזואליזציה זו הינו איש עסקים המעוניין להשקיע/להקים/לקנות מלון בהודו. הוא כמובן ירצה לראות היכן הכי כדאי לעשות זאת- באיזה אזור נמצאים המלונות הפופולאריים ביותר והיכן אין מספיק היצע, ויכול להפיק מידע כגון הקשר בין מיקום המלון (מבחינת נגישות, דירוג האזור, נופ וקרבה למקומות מרכזיים) לדירוג הלקוחות את המלון או לדירוג המיקום של המלון, האם כדאי להקים מלון קרוב לנמל התעופה או ליד נחל או על הר, אילו פרמטרים כגון ניקיון, מיקום וסוג הנכס הכי רלוונטיים בהסתכלות על דירוג המשתמשים ועוד.

כיום, לשני סוגי משתמשים אלו אין מענה מספיק טוב כיוון שלא ניתן לקבל את כל המידע הזה בצורה שניתנת להשוואה ולא ניתן לקבל מידע רוחבי מקריאת טבלה בעלת 20,000 רשומות. בנוסף, לא ניתן להסיק מהטבלה מידע רב שניתן בויזואליזציה כגון מיקום וסביבה. אתרי מלונות המציעים דילים לא מסננים בצורה קפדנית כל כך את המלונות לפי העדפות המשתמש ולא נותנים אופציה לראות את כל הנתונים בצורה מסודרת האחד ליד השני לפי אזור. בנוסף, לרוב כאשר מתקבל מידע על מלון מסוים יש לחפש את האזור במנוע חיפוש נוסף על מנת לבחון את קרבת המלון למקומות מרכזיים ויש לחקור היטב כדי לגלות האם האזור של המלון נחשב טוב או לא, האם הוא מסוכן והאם נעים וכדומה. אני נותנת מענה לכל אלה בויזואליזציה זו המשלבת בתוכה הרבה מידע בצורה אינטואיטיבית, מיידית ומותאמת למשתמש.

## User task

המשתמש: מטייל המחפש מקום לינה בהודו

משימות המשתמש:

- לגלות לאיזה מלון מבין המלונות שמתאימים להעדפותיו דירוג לקוחות גבוה/נמוך יותר.
- לגלות לאיזה מלון מבין המלונות שמתאימים להעדפותיו דירוג המיקום גבוה/נמוך יותר.
- לבצע השוואה בין מלונות שונים באזור בו המטייל מתעניין ומתאימים לסינון שלו.
- לזהות את המלון המתאים ביותר לצרכי המשתמש לפי העדפותיו.
- לזהות את המלונות הנקיים ביותר בתחום הסינון של המשתמש לפי סדר.
- לזהות את קרבת כל מלון המתאים לסינון המשתמש למקומות מרכזיים כמו קניונים, אגמים ונמלי תעופה.
- לזהות את נגישות כל מלון המתאים לסינון המשתמש מבחינת הכבישים המובילים אליו.
- לזהות את אופי סביבת כל מלון המתאים לסינון המשתמש מבחינת תוואי שטח.

## המשתמש: איש עסקים המחפש להשקיע/לקנות/לבנות מלון בהודו

### משימות המשתמש:

- לזהות את האזורים בהם ריכוז המלונות גבוה/נמוך בהתאם לסינון שהמשתמש מכוון אליו (על מנת לדעת היכן חסר היצע ושם כדאי להקים את המלון).
- לזהות את האזורים בהם דירוג המיקום גבוה/נמוך בהתאם לסינון שהוא מכוון אליו (על מנת לדעת איזה איזור הכי כדאי להשקיע).
- לזהות קשרים בין דירוג המיקום לדירוג הלקוחות הכללי בהתאם לסינון שהוא מכוון אליו (קשר בין הרדיוס לצבע. על מנת לדעת את רמת השפעת המיקום על שביעות רצון הלקוחות ועד כמה כדאי להשקיע בנושא זה).
- לזהות קשרים בין הניקיון לבין דירוג הלקוחות הכללי (האם המלונות בטבלה מופיעים במפה עם רדיוס גדול. זאת על מנת לדעת עד כמה ההשקעה בניקיון משפיעה על שביעות רצון הלקוחות).
- לזהות קשרים בין תוואי השטח בסביבת המלון לבין הדירוג שלו (האם לדוגמה ליד הרים מופיעים עיגולים ברדיוסים גדולים יחסית, על מנת לזהות מקומות אסטרטגיים למלון).
- לזהות קשרים בין סוג הנכס לבין דירוג הלקוחות.
- לזהות קשרים בין קרבה לנמלי תעופה ונגישות לבין דירוג המיקום (לדוגמה האם בסביבת נמלי התעופה צבעי העיגולים ירוקים יותר).
- לזהות קשרים בין קרבה לנמלי תעופה ונגישות לבין דירוג הלקוחות הכללי (רדיוסים גדולים בקרבתם).

על מנת להצליח לענות על מטרות אלו ביצעתי מספר אגרגציות ומניפולציות על הנתונים, והן מפורטות תחת פרק "ניקיון ועיבוד הנתונים".



# How

## תהליך בניית הויזואליזציה

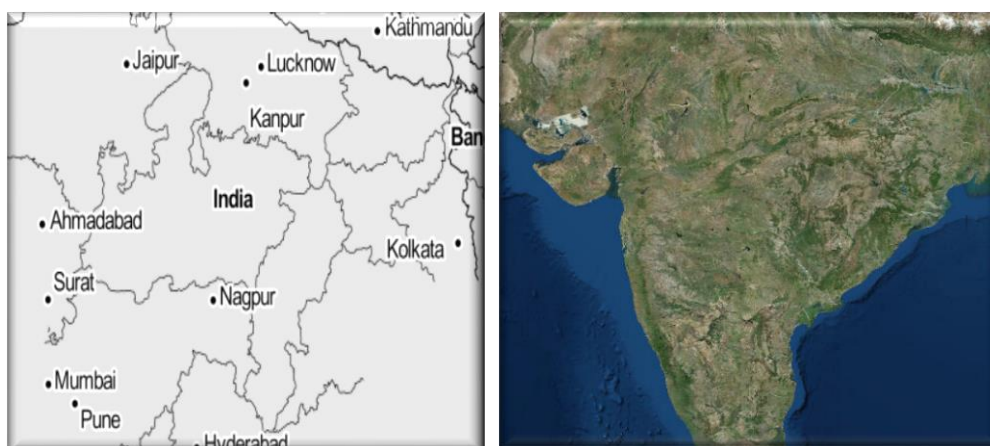
בסיס נתונים המכיל מידע אודות מלונות ברחבי הודו גרם לכך שאחיליט באופן טבעי להשתמש במפה ככלי הויזואליזציה שלי. המפה מוסיפה ערך מוסף, בעזרת ארבעת הממדים הויזואליים שהיא מספקת ועליהם אפרט בהמשך בהרחבה- אורך, רוחב, גודל וצבע, בהשוואה לplot רגיל. כל Plot מסוג אחר חסר את המימד המרחבי שיש למפה ובנוסף אטרקטיבי לעין פחות ממפה, ולכן החלטתי להשתמש במפה ולמקם את התחנות במרחב. נעשה שימוש בחבילת R Leaflet בכדי להציג את מיקום התחנות על המפה.

על מנת לבחור מפה שרלוונטית ספציפית לסוגי המשתמשים שאני מכוונת את הויזואליזציה שלי לשירותם, שיניתי את השכבה של התמונה לגרסה שתתאים לי יותר. לאחר בחינת אופציות רבות ומגוונות לסוגי מפות (ביניהן- HERE.hybridDay, Hydda.Full, Stamen.TonerHybrid, providers\$Esri.WorldStreetMap, CartoDB.Positron, Stamen.TonerHybrid, providers\$Esri.WorldImagery, providers\$Esri.NatGeoWorldMap) בחרתי באופן ספציפי ומחושב בשכבת- providers\$Esri.NatGeoWorldMap אשר עוזרת לענות על ה-user tasks שהצגתי. בעזרת שכבה זו, ניתן לראות את נמלי התעופה, תוואי השטח, אגמים, נחלים, כבישים, הרים, מדבריות, מדשאות ואת הגובה של כל אזור – דברים שחשובים כאשר בוחרים מקום לינה במדינה בה רוצים לטייל או לחילופין כשרוצים לקנות/לבנות מקום לינה למטיילים. הנגישות של המלונות מוצגת במפה בצורה כזו שניתן לראות את הקרבה לשדות התעופה ואת כל הכבישים שעוברים באזור המלונות. הנוף מהמלון ניתן להסקה בכך שניתן לראות האם המלון נמצא באזור עירוני, עמוס, הררי, גבוה, נמוך, בקרבת נהר או אגם וכדומה. כל אלו מושגים באמצעות המפה שבחרתי, לדוגמה באזור עירוני אל מול טבעי המפה תראה כך:



איור 1 - המפה הנבחרת, אזור עירוני מול הררי

דוגמה לשתי מפות שפסלתי:



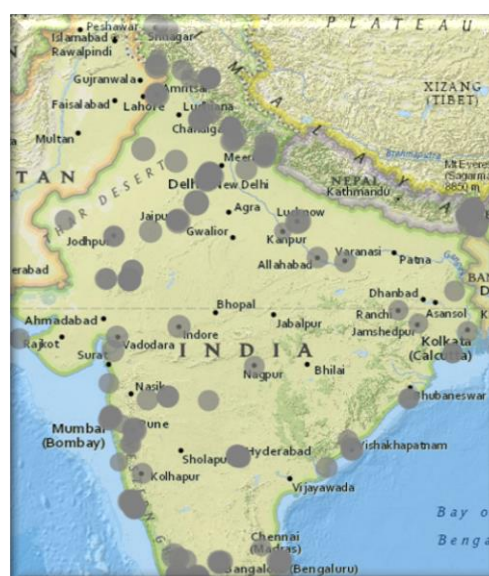
איור 2 - מפות שלא נבחרו

לאחר בחירת המפה, יצרתי עליה עיגולים על מנת שאוכל לייצג על ידי הסמן עוד אינפורמציה ולא רק מיקום. החלטתי להעזר בעיגולים על מנת להציג 2 סוגי מידע-

האחד, דירוג הלקוחות את המלון. בבסיס הנתונים היה ערך נתון לכל מלון, המציין את כמות הלקוחות אשר דירגו את הערכת המלון כ- Excellent, Very Good, Average, Poor, Terrible. לפי כמויות המדרגים לכל הערכה נעשה ממוצע משוקלל שהוא דירוג הלקוחות בו אשתמש לכל מלון. לפי ראות עיניי זהו אחד הפרמטרים החשובים ביותר למימוש מטרת ויזואליזציה זו ועל כן החלטתי לייצג אותו באופן בולט וניכר – רדיוס העיגולים.

ארצה לבחור רדיוס שלא יהיה מאוד קטן ויגרום להעלמות העיגולים גם אם דירוג מלונות אלו הוא נמוך, ולכן החישוב נעשה באמצעות הנוסחה:

$$radius = \begin{cases} 5, & mmt\_review\_score < 2 \\ mmt\_review\_score * 2.5, & else \end{cases}$$



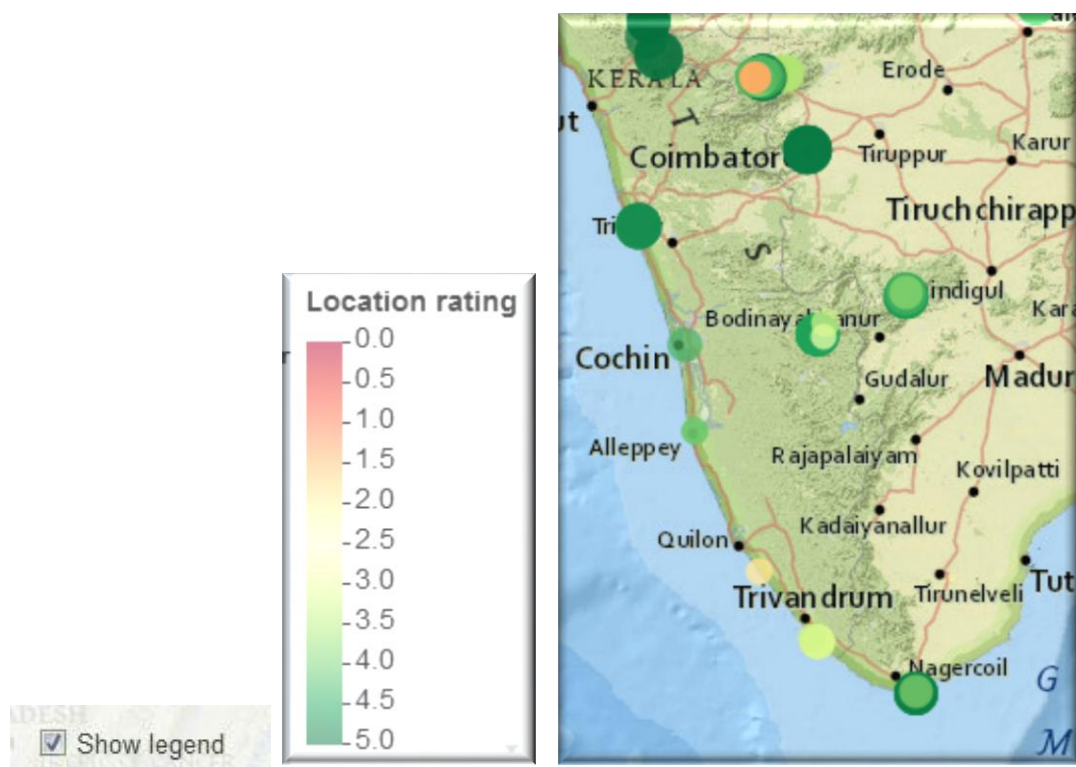
איור 3 - יצירת העיגולים על המפה

השני, מאפיין נוסף שהיה חשוב לי לשים עליו דגש והוא דירוג המיקום, שחשוב מאוד במיוחד בהודו בה ההבדלים בין האזורים משמעותיים וישנם אזורים מסוכנים מאוד, ורלוונטי גם למטיילים המחפשים מקום לינה וגם לבעלי המלונות העתידיים. עבור תצוגת פרמטר זה השתמשתי בתכונה `traveller_rating` שזהה בערכה לתכונה המיקום שהפרדתי והוצאתי מתוך `traveller_rating` תחת `Location`. בחרתי להציג את טווח הערכים הרציף של התכונה (מ-0 עד 5) בעזרת סקאלת צבעים שנעה מירוק (הדירוג הכי טוב) עד אדום (הדירוג הכי גרוע) דרך צהוב על מנת להפוך את הקריאה לאינטואיטיבית. הנקודות בשקיפות יחסית גבוהה על מנת שעדיין ניתן יהיה לראות את המפה מאחוריהן.

בתחילה, רציתי שצבע הנקודות ייצג את "סוג המטייל" שהמלון מתאים לו (משפחות, זוגות, עסקים...). באופן בדיד, אך לאחר ניסוי ומחשבה הסקתי כי אין בכך טעם מכיוון שמי שמחפש מלון לעסקים אין סיבה שיראה מלונות לקבוצות חברים ויסרוק את כל המפה בחיפוש אחר צבע של מקומות לעסקים, ובנוסף מכיוון שיש הרבה מאוד מלונות בבסיס הנתונים שמוגדרים למספר סוגי מטיילים (ורק כמות המצביעים לכל סוג שונה) אין בכך הרבה הגיון. זו הסיבה שבחרתי בסופו של דבר לייצג את דירוג המיקום כצבע העיגול ולבצע סינון על פי סוג המטיילים במלון כך שהמשתמש יראה רק את המלונות שרלוונטיים עבורו.

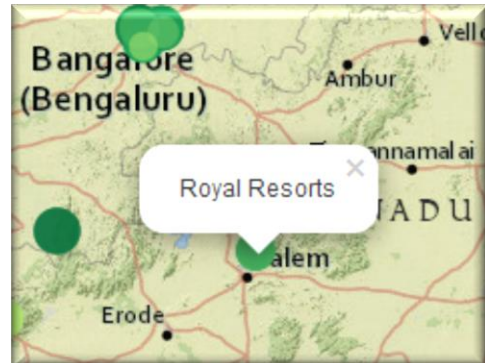
בנוסף, הוספתי מקרא עם סקאלת הצבעים ו**כותרת** על מנת שידעו מיידית מהיכן נובעים השינויים בצבעים בין העיגולים ומה הם מסמלים.

מאוחר יותר הכנסתי גם אופציה להסתיר את הסקאלה מכיוון שהציון עצמו של דירוג המיקום פחות קריטי מאשר קבלת הערכה על פי הצבעים, ולכן לאחר מבט והפנמה של מה הצבעים מייצגים פחות חשוב בעיניי שהסקאלה תהייה מול עיני המתבונן לאורך כל הדרך. בצורה כזו, המסך יהיה פחות עמוס.



איור 4 - צבעי העיגולים והסקאלה שלהם במפה

בשלב הבא, על מנת לדעת את שם המלון שהמשתמש מתעניין בו הוספתי אופציה ללחיצה על כל עיגול והצגה של שם המלון (PopUp).



איור 5 - שם המלון בלחיצה על העיגול

על מנת להציג מידע רלוונטי בלבד ולקצר את משך סריקת המידע, בחרתי לתת למשתמש שליטה על בחירת רמת המלון שהוא מחפש (מבחינת כמות כוכבים). כמות הכוכבים היא דירוג אובייקטיבי יותר מאשר דירוג הלקוחות של המלון שהוא דירוג סובייקטיבי ולכן מעניין לראות את שני הפרמטרים הללו. בתחילה, הצגתי את כמות הכוכבים כטבלה נגללת ולאחר מחשבה הבנתי שמרבית האנשים יעדיפו לראות טווח של ערכים בקטגוריה זו ולבחור מביניהם כי בדרך כלל הדרישה לכמות כוכבים מסוימת לא חד משמעית ללקוחות בחיפוש מלון. לכן העברתי את בחירת כמות הכוכבים לבחירה על סקאלה רציפה שמאפשרת טווחים.

בתחילת העבודה גם ניסיתי לעשות slider לכל אחד מהערכים שהוצאתי מתוך פרמטר Location, Hospitality, Facilities, Cleanliness, Value for Money, traveler\_rating כלומר Food אך לאחר שראיתי כי הדבר יוצר עומס רב על הויזואליזציה בחרתי להציג רק את מה שנראה בעיניי הכי חשוב והוא התמורה לכסף (כלומר, עד כמה פרמטר זה קריטי עבורו 0 ועד 5). גם בחירה זו השתנתה בהמשך כמוסבר בהמשך. את דירוג המיקום והניקיון הצגתי בדרכים אחרות- דירוג המיקום בעזרת סקאלת הצבעים כפי שנאמר כבר ודירוג הניקיון יוסבר בהמשך. בנוסף, בתחילה בחרתי ב slider חד ערכי לתמורה לכסף, ולאחר מכן הבנתי שבחירת המשתמש בתחום זה לא תהיה חד משמעית כל כך להנחתי, לכן החלטתי לתת למשתמש לבחור טווח ערכים לתמורה לכסף שתיתן הרבה יותר תוצאות מתאימות מאשר אם הבחירה הייתה לערך יחיד ועל המשתמש היה לבצע הרבה יותר חיפושים כדי להגיע לתוצאה מספקת. לאחר מחשבה הבנתי כי ערך התמורה לכסף שלקוח מחפש תמיד יהיה הגבוה ביותר (לא מדובר כאן במחיר אלא בתמורה) ולכן אין טעם לתת לו בחירה של ערך זה ולכן הורדתי בחירה זו.

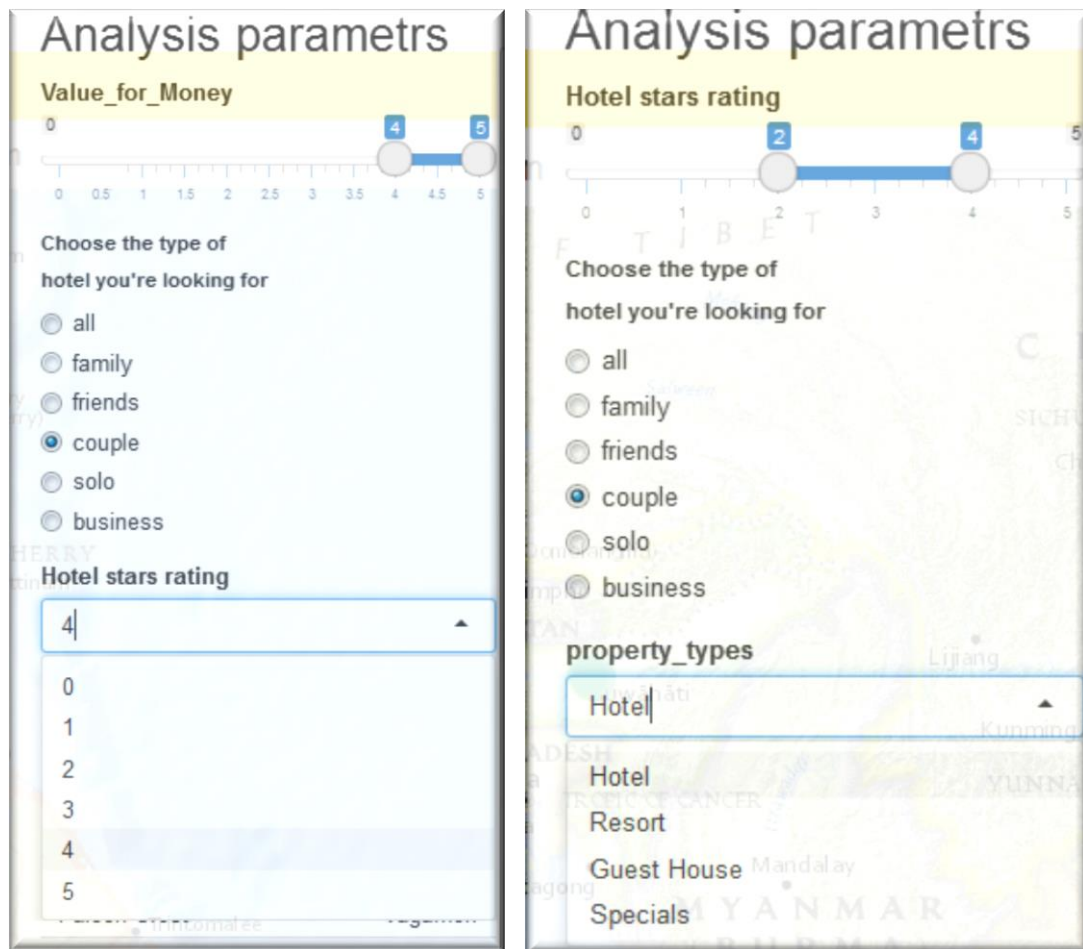
נתתי למשתמש גם שליטה על בחירת סוג המלון שהוא מחפש- האם מלון למשפחה, לזוג, לעסקים, לחברים או לבודד בעזרת כפתורי "רדיו".

בנוסף, לאחר שהחלפתי את השליטה על מספר הכוכבים מטבלה נגללת בדידה לסקאלה רציפה שמאפשרת טווחים בחרתי לתת למשתמש שליטה בטבלה הנגללת על סוג הנכס שהוא מחפש- מלון, בית הארחה, בית סירה וכו'. לפרמטר זה יש משמעות והוא בדרך כלל חד משמעי אצל שני סוגי המשתמשים בויזואליזציה זו.

בכל פעם שהמשתמש משנה את בחירתו מופיעים מלוות שונים (באמצעות subset ומניפולציות יצרתי בסיס נתונים חדש בכל פעם בהתאם לבחירות המשתמש), הרדיוס מחושב מחדש, והטבלה



שמוסברת בהמשך מוצגת מחדש עם מלונות מתאימים לסינון. שליטת המשתמש בפרמטרים הנ"ל מאפשרת לויזואליזציה להיות רלוונטית ויעילה ומאפשרת למשתמש to explore, locate, browse the data.



איור 6 - ממשק המשתמש לפני ואחרי החלפת הפרמטרים

בנוסף, כמטיל בהודו אחד הדברים הקריטיים ביותר, שאינם מקבלים חשיבות מספקת בתרבות ההודית כמו בתרבות המערב, הוא ניקיון מקום הלינה. לכן, החלטתי שברצוני לתת גם לאספקט זה מקום בויזואליזציה שלי. השתמשתי ב `traveler_rating` ומתוכו משכתי את דירוג הלקוחות ל- cleanliness. מכיוון שלא רציתי ליצור עוד עומס על המפה בחרתי להציג טבלה לצד המפה, שתספק מידע על חמשת המלונות הנקיים ביותר בהתאם לסינון המשתמש. בעזרת תכנת R בחינת ה cleanliness מתבצעת בכל פעם שהמשתמש משנה את בחירתו, ובהתאם נלקחים חמשת המלונות עם הדירוג הגבוה ביותר. על מנת לעזור למשתמש למצוא את המלון המתאים מהטבלה בתוך המפה על מנת לקבל עליו את המידע הנוסף שהמפה מציעה כגון תוואי השטח, נגישות, רדיוס העיגול המצביע על דירוג המשתמש וצבע העיגול המצביע על דירוג המיקום, הוספתי לטבלה גם את העיר בה נמצא המלון, וכך המשתמש יכול לבצע תקריב של המפה לעיר הספציפית ולמצוא בה את המלון. את הטבלה שילבתי בתצוגת shiny.

בתחילה הוספתי ליד שם כל מלון את הכתובת המדויקת שלו, אבל הדבר יצר עומס מידע ויצר מצב שעל מנת לראות את כל חמשת המקומות היה צריך לגלול (או להזיז) את הטבלה למעלה ולמטה. הכתובות הנתונות בבסיס הנתונים ארוכות מאוד ולא אחידות במידע שהן מספקות. כאשר רואים רק את העיר, ניתן לראות את כל הטבלה בהיבט יחיד במסך, וניתן להסיק בצורה טובה היכן המלון

ממוקם. בכל אופן תייר שרואה את הכתובת המלאה לא מקבל אינפורמציה רלוונטית ולרוב שם העיר מספיק.

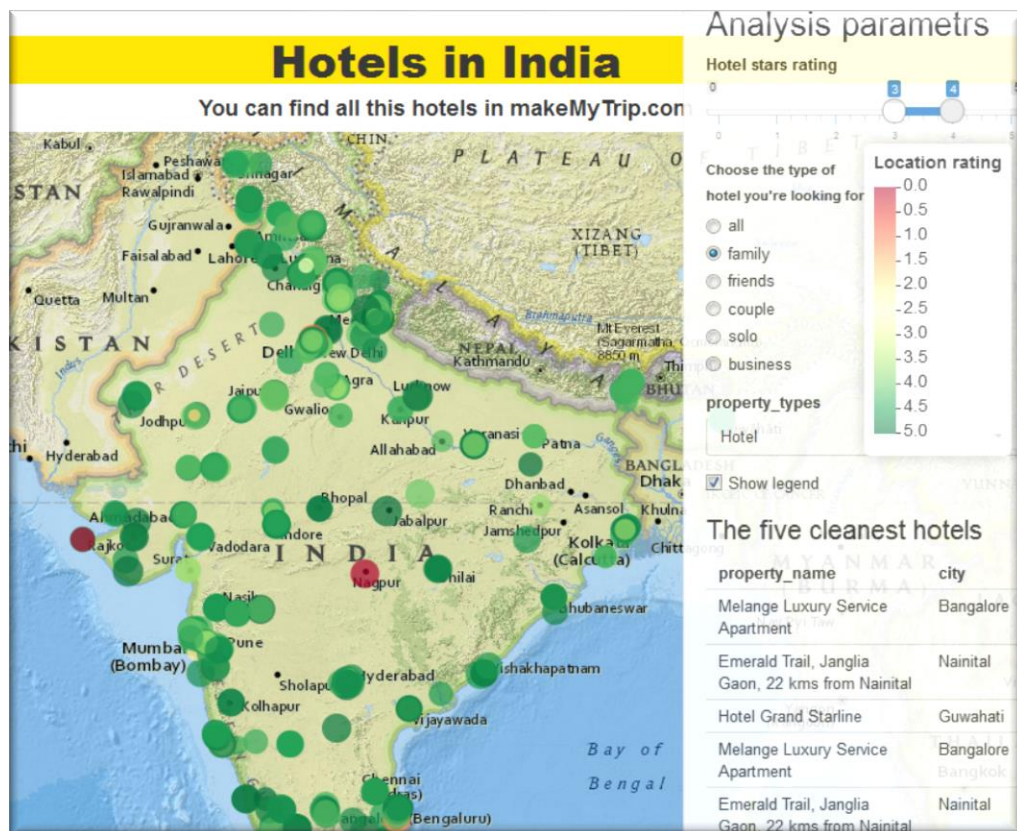
property_name	city
Hotel Sea Shell - Port Blair	Port Blair
Falcon Crest	Vagamon
Le Sutra - The Indian Art Hotel	Mumbai
Fog Hills Manali	Manali
Hotel Harmony Inn	Meerut

איור 7 - טבלת חמשת המלונות הנקיים ביותר בהתאם לסינון המשתמש

### Visual encoding

מיקום המלון	מיקום על המפה לפי lng, lat
דירוג מיקום המלון	צבע העיגול
דירוג הלקוחות את המלון	רדיוס העיגול
שם המלון	"ערך" העיגול – המופיע בלחיצה (PopUp)
רמת ניקיון מלון היחסית	סדר בטבלת המלונות הנקיים ביותר
מיקום המלונות הנקיים	העיר בטבלת המלונות הנקיים ביותר
סוג הנכס (מלון, חדר הארחה...)	לפי בחירת המשתמש בטבלה נגללת
סיווג סוג הלקוחות המתאימים למלון	לפי בחירת המשתמש – כפתורי "רדיו"

## תוצאות הוויזואליזציה



איור 8 - תוצר סופי

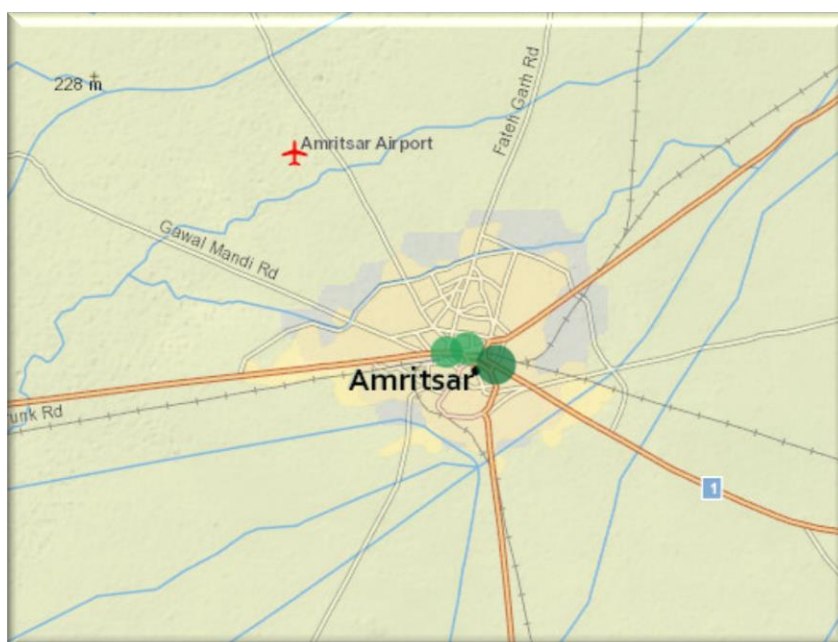
דוגמה נוספת (כולל טקסט):



איור 9 - דוגמת תוצר סופי עם טקסט

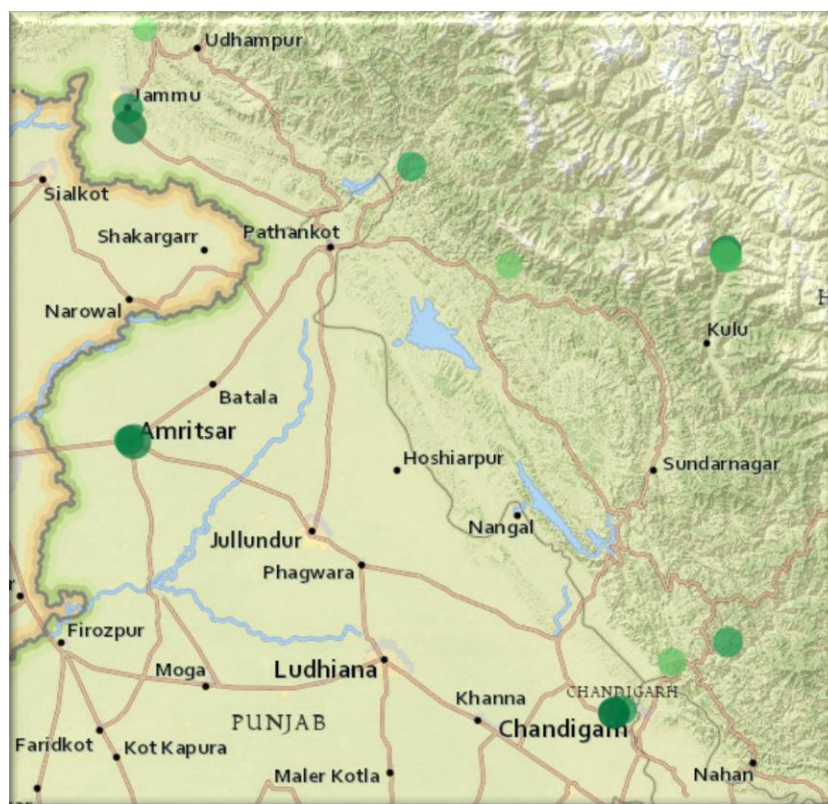


בתקריב:



איור 10 - תוצר סופי בתקריב – אזור עירוני

אזור חוץ עירוני:



איור 11 - תוצר סופי בתקריב- אזור חוץ עירוני

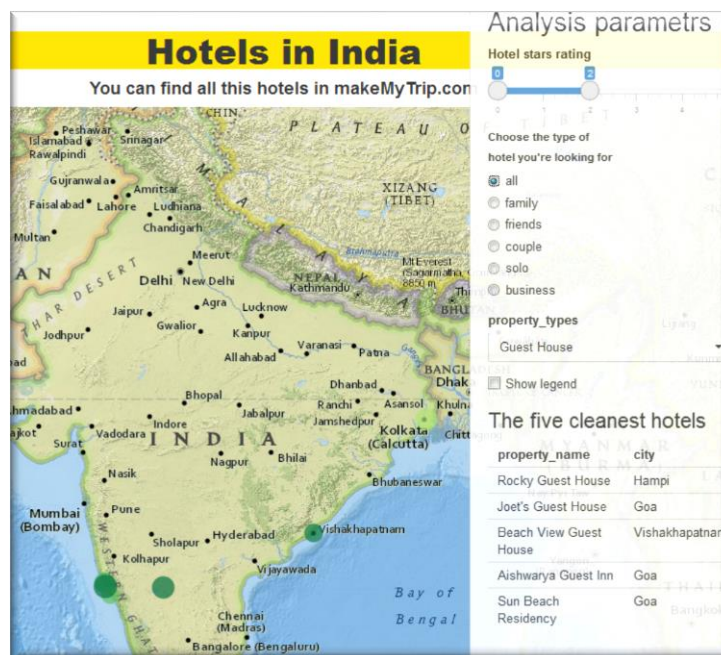


## אינטראקטיביות הויזואליזציה

לצורך אינטראקטיביות הויזואליזציה נעשה שימוש בחבילת shiny והוחלט לתת למשתמש אפשרות לסנן את המלונות המוצגים ובכך גם ליצור טבלת חמישה מלונות מובילים בניקיון מותאמת, בהתאם לטווח הנבחר של מספר הכוכבים, התאמה לסוג מטייל מסוים וסוג הנכס שהמשתמש מחפש.

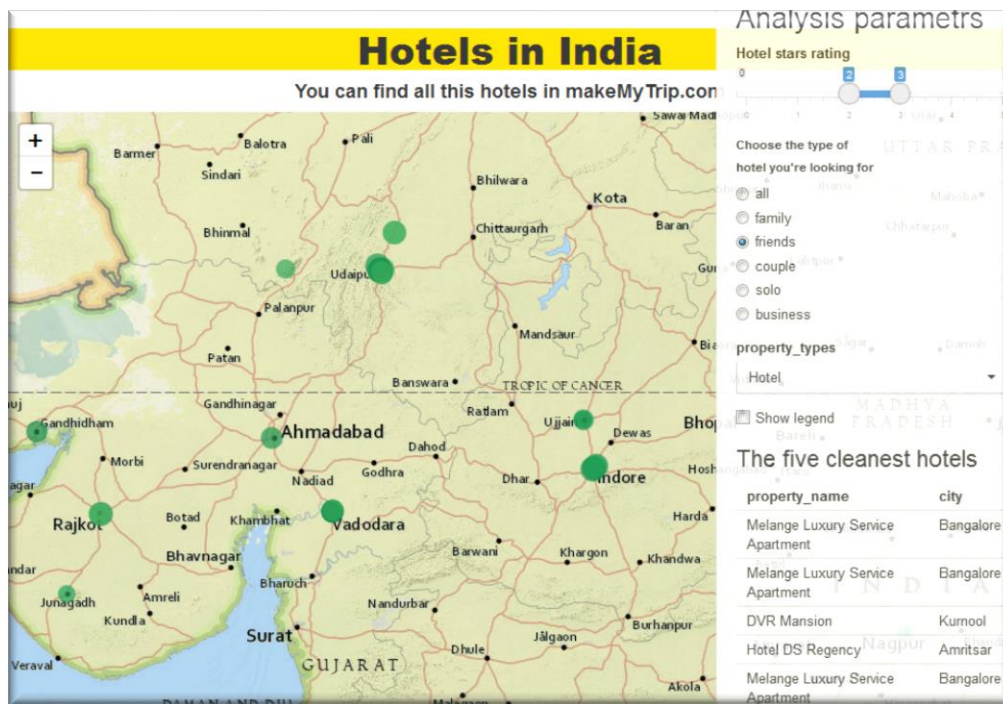
איור 12 - ממשק המשתמש

שינוי כלל הערכים (טווח מספר הכוכבים, סוג המטיילים וסוג הנכס המבוקש) – להשוואה מול- איור 8 - תוצר סופי :



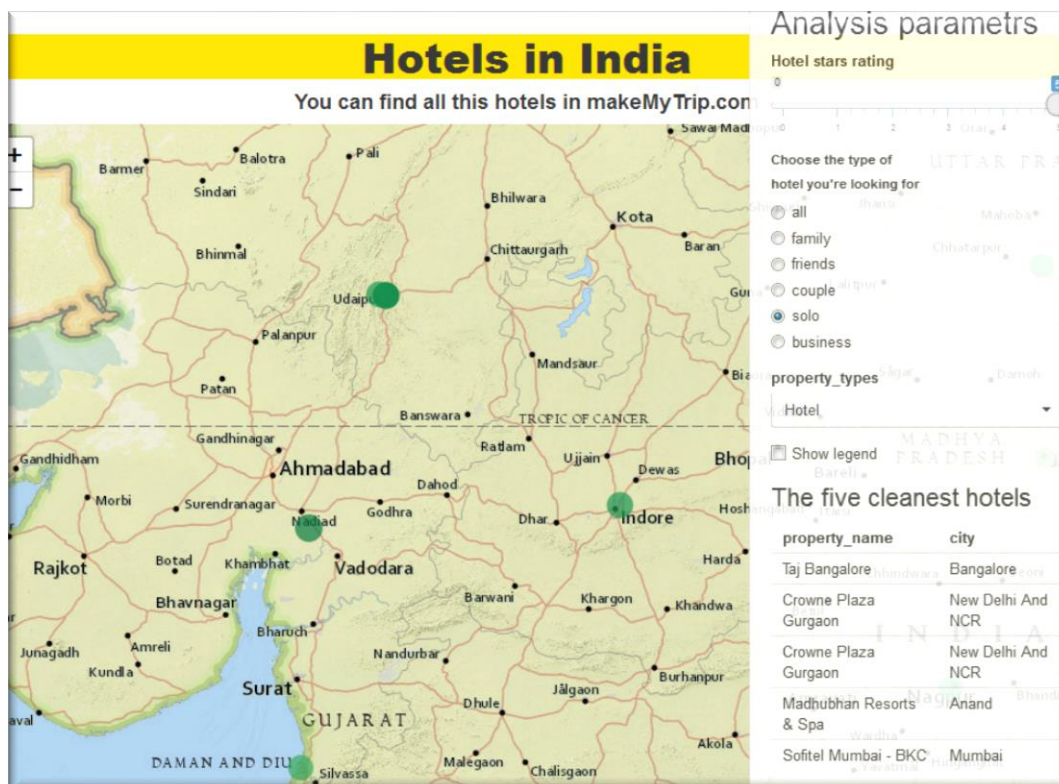
איור 13 - תוצר סופי עם שינוי ערכי הסינון

שינוי נוסף בכלל הערכים בתקריב:



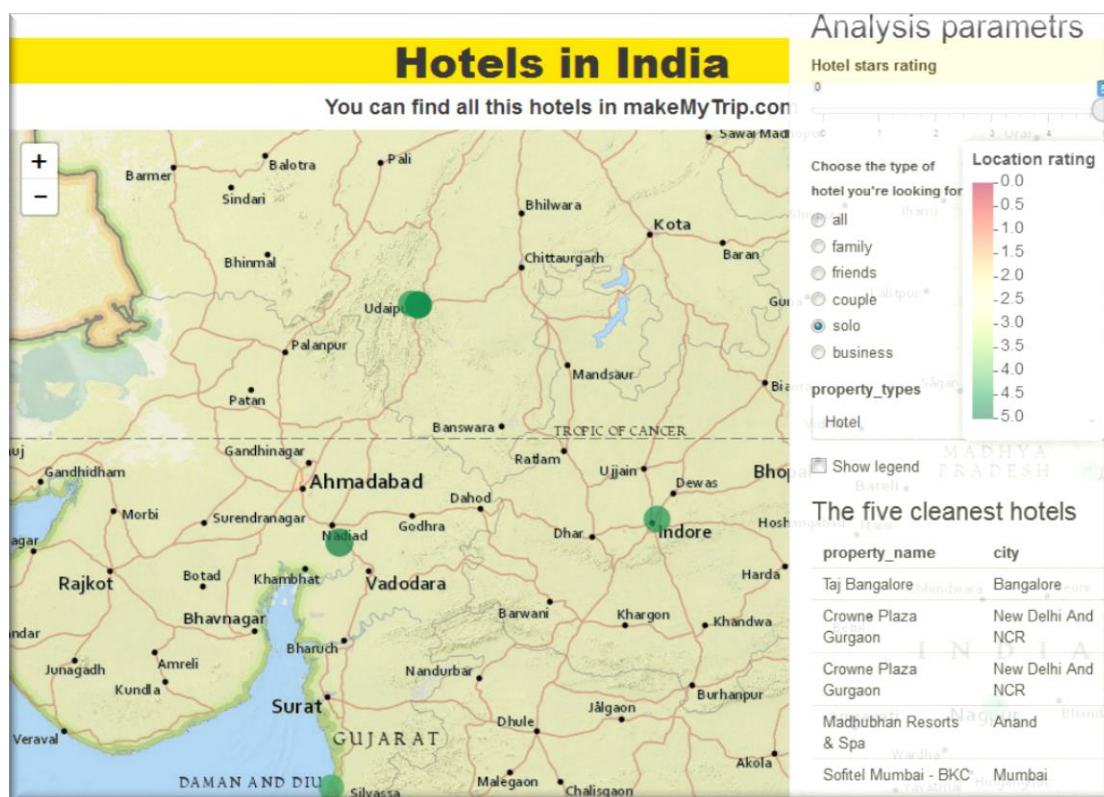
איור 14 - שינוי נוסף בערכי הסינון

שינוי הערכים באותו התקריב ובאותו האזור – להשוואה מול איור 14 - שינוי נוסף בערכי הסינון :



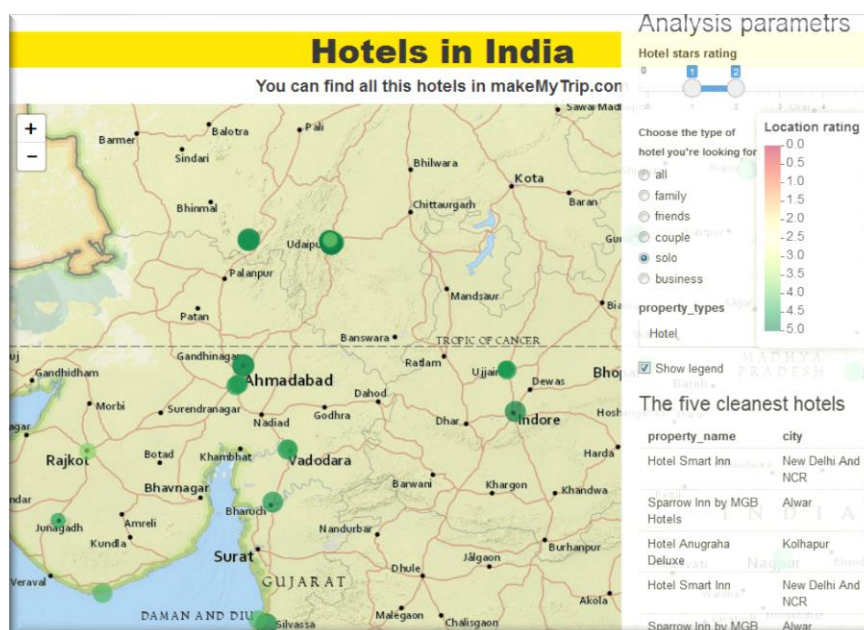
איור 15 - תוצר סופי נוסף

אותם ערכים עם מקרא:



איור 16 - הוספת מקרא

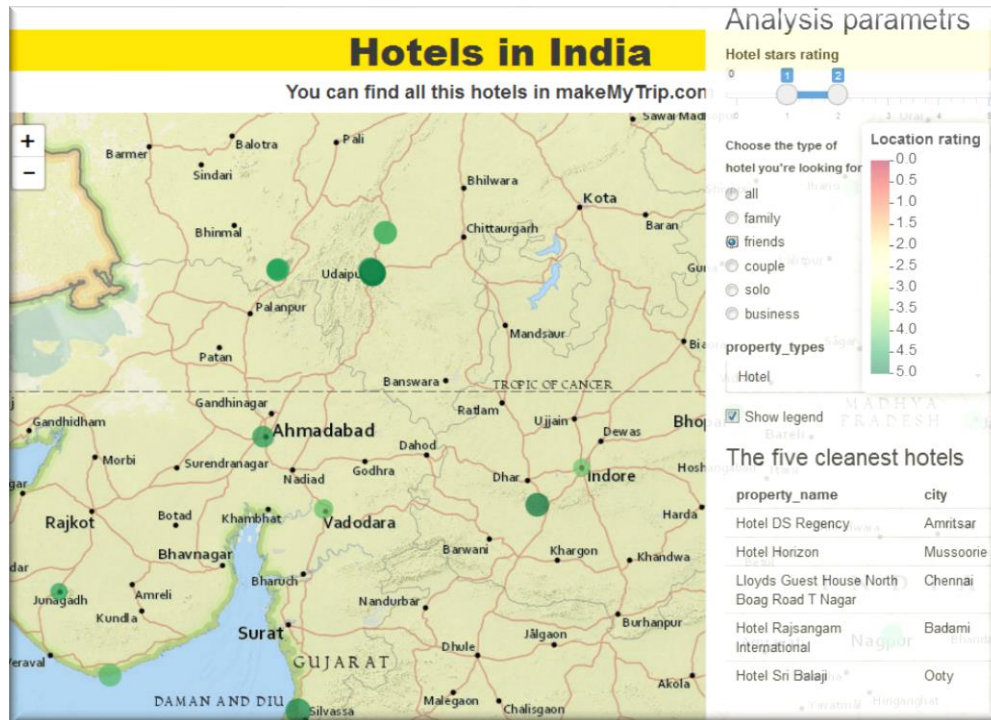
שינוי בכמות הכוכבים בלבד:



איור 17 - שינוי בטווח מספר הכוכבים

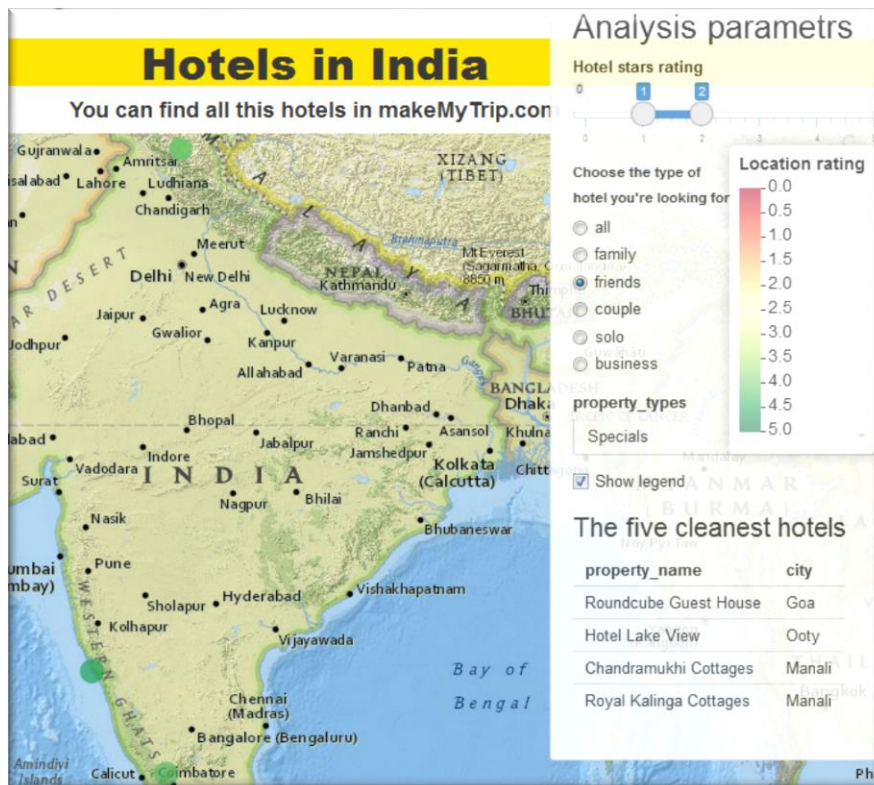


שינוי נוסף, קעת בסוג המטייל בלבד:



איור 18 - שינוי בסוג המטייל

שינוי נוסף, קעת בסוג הנכס בלבד:



איור 19 - שינוי בסוג הנכס

# Value of visualization

## יתרונות

- העובדה כי הויזואליזציה מאגדת מידע רב ומציגה אותו בתצוגה אחת חוסכת זמן למשתמש.
- הויזואליזציה מושכת במבט ראשון ונעימה לעין.
- תפעול הויזואליזציה אינטואיטיבי מאוד למשתמש – תקריב בעזרת העכבר (או סמני פלוס ומינוס על המפה), עזרי סינון אינטואיטיביים (טווח למספר הכוכבים, טבלה נגללת לסוג הנכס וכפתורי רדיו לסוג המטייל). צבעי העיגולים אינטואיטיביים (ירוק כטוב ואדום רע).
- בנוסף, כותרות המסבירות את תפקיד כל חלק בממשק המשתמש.
- המפה משמעותית ונוחה לשימוש להצגת מיקום המלונות, הנגישות, הקרבה למקומות מרכזיים ותוואי השטח סביב הנכס.
- נוח לראות ערכים קיצוניים שמתאימים מאוד או לא מתאימים כלל למשתמש (צבע אדום, עיגול קטן או גדול במיוחד, התמקדות באזור ספציפי במפה).
- ניתן להסיק על מגמות לדוגמה- עיגול ירוק יותר בדרך כלל גדול יותר – מלונות במיקומים טובים יותר מקבלים בדרך כלל דירוג לקוחות טוב יותר.
- ניתן להשתמש בויזואליזציה למגוון user tasks ולכמה סוגי משתמשים.
- ניתן להסיק מסקנות בצורה מהירה יחסית לגבי אזורים כדאיים ללינה ולהשקעה.
- הויזואליזציה כפי שיוצרה בעזרת הממשק האינטראקטיבי shiny מאפשרת למשתמש חופש פעולה ומרווח תמרון בבחירת המידע שברצונו לקבל- הויזואליזציה נותנת חופש למשתמש לבחור את סוג המטייל, מספר הכוכבים (גם כטווח) ואת סוג הנכס שבו הוא מעוניין.
- ניתן לקבל אופציות מותאמות בצורה טובה ומצומצמת להעדפות המשתמש. הויזואליזציה מבצעת הבחנה בין עיקר לתפל. הויזואליזציה מותאמת לuser tasks בהכונה מירבית.
- זמן הריצה של האלגוריתם מיידי ומספק חווית משתמש טובה.
- אין צורך בידע קודם על הודו, המלונות או האזורים בה מה שהופך את הויזואליזציה ליעילה עבור כל משתמש.

## חסרונות

- קשה מאוד להסיק מסקנות חד משמעיות מהויזואליזציה לגבי מגמות והיא למעשה נשארת רק בגדר מידע אשר מרמז על המגמות אך לא יכול להצביע עליהן בבירור.
- לא ברור לכולם בצורה אינטואיטיבית מה פשר ההפרש בגדלים של העיגולים ואין על כך הסבר מילולי בויזואליזציה.
- הכלי לא ספציפי מספיק – אין מידע קונקרטי על המלונות כגון כמות הכוכבים של המלון הספציפי אם בחרת טווח ערכים. גם מבחינת הצגת ערכים מספריים ספציפיים (גם לדוגמה ברמת הניקיון, מוצגים שמות חמשת המלונות הנקיים ביותר אך ייתכן כי המלון הכי נקי גם הוא מלוכלך ולא ניתן לדעת זאת) – יכול להטעות. בתחילה שמתי את ציוני הניקיון בטבלה אך הדבר יצר עומס מידע ונראות לא טובה של הויזואליזציה ובנוסף, כמעט תמיד הציון היה 5 מה שהפך את הדבר למיותר.
- בעקבות הנקודה הקודמת, לא ניתן לזהות קשרים ממשיים בין כמות הכוכבים של מלון לבין ניקיון המלון לדוגמה מכיוון שאין דרך לדעת לא את כמות הכוכבים הספציפית ולא את ניקוד הניקיון הספציפי.
- קשה לזהות קשרים בין כמות הכוכבים לדירוג הכללי (ניתן לראות האם כשמשנים את כמות הכוכבים העיגולים המופיעים קטנים או גדולים יותר באופן כללי, אך קשה להשוות בין עיגולים שונים כשהם כבר לא על המסך). באופן כללי, המשתמש יתקשה להשוות פרמטר זה (דירוג הלקוחות המתבטא ברדיוסי העיגולים) כשישנה פרמטרים בבחירתו.
- הויזואליזציה אינה ידידותית לעיוור צבעים – Colorblind friendly.

- לרוב המלונות ציוני מיקום יחסית טובים (מעל 2.5) ולכן רובם בגווי ירוק וקשה להבדיל ביניהם.
- במקרה בו ישנם מלונות רבים באותו המקום- הנקודות עולות האחת על השנייה וקשה מאוד לזהות ולקבל popup של הנקודות התחתונות / הקטנות יותר.
- במקרה בו מספר נקודות נמצאות באותו האזור הן לפעמים מסתירות את שם המקום על המפה.
- קשה לזהות קשרים בין כמות הכוכבים לסוג הלקוחות לסוג המלון. ניתן לראות אם מופיעות יותר או פחות אופציות למלונות כשמשנים רק גורם יחיד, אך לא ניתן לזהות את המגמות בצורה מדויקת כיוון שאין נתונים מספריים של כמות הכוכבים אם בוחרים בטווח ערכים.

## הצעות לשיפור ושדרוג הוויזואליזציה

- הוספת מידע נוסף למפה – מידע ספציפי כמו כמות כוכבים למלון, דירוג מספרי על ידי הלקוחות של המלון, דירוג הניקיון, כתובת מדויקת ועוד.
- אינפורמציה הקטן יכולה להיות יותר אינפורמטיבית.
- אפשר להשתמש בצבע לשימוש אחר ולתת דירוג מיקום לאיזור ולא פר מלון (יש צורך לעשות עיבוד מתקדם של הנתונים בבסיס הנתונים לשם כך על מנת להוציא דירוג לאיזור מסוים).
- ליצור אופציות להשוואה בין מלונות ספציפיים שנבחרים – השוואה לפי דירוג לקוחות לדוגמה תפתור את בעיית השוואת רדיוסי העיגולים.
- הוספת פרמטרים נוספים לניתוח הנתונים.
- היה יותר רלוונטי אם הייתי משיגה מידע על המחיר של כל מלון ולא רק על "התמורה למחיר". כדאי להרחיב את בסיס הנתונים ולהכניס לתוכו מחיר ממוצע לחדר במלון ואז להציג את המידע בויזואליזציה. זה מידע מאוד מרכזי למשתמשים בויזואליזציה זו.
- כדאי היה לקשר את הוויזואליזציה לאתר האינטרנט של הזמנת חדרים במלון ואז היה ניתן לסנן גם לפי תאריכים בהם יש חדרים פנויים במלון. כדאי להרחיב את בסיס הנתונים בהמשך למען מטרה זו.
- כרגע הקישור בבסיס הנתונים לדפי האינטרנט של האתרים לא עובד. אם היה עובד, הייתי מוסיפה אותו ב popup של כל מלון.
- מומלץ בהמשך להוסיף אופציה לחיפוש ספציפי של מלון לפי שם או כתובת. כך נוכל למצוא במפה את המלונות שבטבלת המלונות הנקיים ביותר והמשתמש יוכל לדוגמה למצוא מלונות שהמליצו לו עליהם ועוד.
- יש להרחיב את בסיס הנתונים על מנת לקבל יותר תוצאות בכל הקטגוריות (במיוחד בחיפושים של resorts או בתי הארחה).

## ערך הוויזואליזציה לפי המדדים שנלמדו בכיתה

- Time - יכולת לחסוך בזמן :  
לדעתי הוויזואליזציה מאוד אינטואיטיבית להבנה וניתן בקלות לעבוד עם ממשק shiny בכדי לנתח את הנתונים ולהסיק מסקנות לגבי המלונות השונים. אני מאמינה שהוויזואליזציה חוסכת זמן לשני סוגי המשתמשים שהצגתי בחיפושיהם אחר תשובות למטרתם. כמו כן ניתן לזהות בקלות את המלונות בעלי הדירוג הגבוה והנמוך יותר ובעלי המיקום הטוב יותר והטוב פחות. ניתן גם לזהות באופן מיידי את האזורים בהם היצע המלונות גדול יותר או נמוך יותר ולזהות מלונות נגישים, קרובים לשדות תעופה, וקרובים לנוף בו המשתמש מעוניין. אולם קשה לבצע השוואה בין דירוגי המלונות שאינם נמצאים בסמוך האחד לשני מכיוון שקשה להשוות בין גדלי הרדיוסים בצורה מספקת ולא ניתן בויזואליזציה מידע מספרי על הדירוג. גם ללא ידע קודם על האזורים בהודו כל משמש יוכל להפיק תועלת ולהבין בזמן קצר את אשר מוצג בויזואליזציה. מכיוון שוויזואליזציה זו מאגדת בתוכה נתונים רבים, אשר קשה מאוד לקרוא מתוך טבלה (לדוגמה מיקום) ולהסיק על בסיסם מסקנות, אין ספק

שהוויזואליזציה חוסכת זמן במענה על שאלות המשתמש. הוויזואליזציה מפרידה בין עיקר ותפל עבור המשתמש ונותנת לו את האפשרות להסיק מסקנות על המלונות שתואמים לו בלבד.

○ Insights - יכולת להבין תובנות חדשות :

המידע המוצג בוויזואליזציה רלוונטי ביותר עבור מטיילים שמחפשים מקום לינה ועבור אנשי עסקים המחפשים מלון להשקעה, קנייה או בנייה. ויזואליזציות המפה מאפשרות להסיק מסקנות שאדם לא יוכל להסיק מהסתכלות על טבלה בעלת 20,000 רשומות בעצמו. את המידע המוצג בוויזואליזציה היה קשה להפיק מהנתונים עצמם, נעשה עיבוד על הנתונים ובנוסף המידע מוצג בצורה וויזואלית הממחישה את מיקום המלונות על המפה, דירוג המיקום שלהם ודירוג הלקוחות שלהם. היא נותנת את האפשרות לסנן את המידע כך שמול המשתמש יעמוד רק מידע רלוונטי לו והדבר יקל עליו בהסקת המסקנות הדרושות לו למענה על ה user tasks שלו.

○ Essence - מהות, סיכום כל הידע והעברתו באופן תמציתי :

המפה מציגה את התמונה הגדולה יותר עבור המידע שלא ניתן להוציא מהנתונים הגולמיים בצורת הצגתם בטבלה. החלוקה לאזורים בהודו – האזורים המתוירים יותר ופחות, הבטוחים יותר ופחות, נותן אופציה להשוות מבחינת מלונות המתאימים לסינון המשתמש, מבחינת ניקיון, דירוג המיקום ודירוג הלקוחות. ההשוואה יחסית קשה לביצוע מבחינת גדלי העיגולים, אך עדיין המידע שניתן להפיק מהוויזואליזציה הוא רב. החלק הלוקה בחסר שהוזכר יכול היה להשתפר אם הייתה תצוגה מספרית של ערכי דירוג הלקוחות והמיקום, ובכלל שיפור הוויזואליזציה היה יכול להתבצע עם הוספת מחירי חדרים ממוצעים לבסיס הנתונים ולוויזואליזציה, קישור להזמנה באתר המלון ואופציה לחיפוש מלון לפי שם או כתובת. הוויזואליזציה אשר יצרתי אכן מכילה תמצית של עיקר הידע ולמעשה מאגדת בתוכה כמות גדולה מאוד של מידע, אשר קשה היה לחקור בצורה אחרת, עם זאת ישנו מידע אשר היה ביד"י ובחירתי שלא לעשות בו שימוש (כדוגמת המידע על סוגי חדרים).

○ Confidence - אופן בתוצאות הוויזואליזציה :

לדעתי באופן כללי הוויזואליזציה משרה ביטחון לגבי אמינות הנתונים. תוצאות הוויזואליזציה אמינות בנוגע למלונות אך לא בנוגע לסוגי נכסים אחרים, זאת מכיוון שאין מספיק מידע בבסיס הנתונים על נכסים שאינם מלונות. לכן, על כל נכס בפני עצמו המידע אמין אך לא ניתן להסיק מסקנות לגבי סוגי הנכסים שאינם מלונות באופן כללי. לגבי המלונות, הוויזואליזציה אמינה באופן כללי אך יש לה בעיות מסוימות בעיניי :

1. ההבדלים בין הרדיוסים של העיגולים לא כל כך משמעותיים לעין. בנוסף, קשה מאוד לבצע השוואה בין גדלי עיגולים שאינם סמוכים האחד לשני ובאותו הגובה במפה. הציון המספרי של דירוג הלקוחות לא מופיע לכל מלון ולכן ניתן להסתמך רק על גודל העיגולים. ניסיון ההשוואה ביניהם יכול להטעות.

2. רדיוס העיגולים בוויזואליזציה אינו מחושב בצורה לנארית אלא באמצעות פונקצייה שמחזירה ערך מינימלי של 5 (על מנת לא לפספס עיגולים קטנים מידי) מה שיכול לגרום להטיה בניתוח הנתונים.

3. בנוגע לדירוג המיקום - אי אפשר לדעת לפי מה דירגו את המיקום- בטיחות, ביטחון, קרבה למקומות בילוי, אזור שקט וכו'. אי אפשר לדעת מי דירג והאם הוא חיפש מה שאותו הלקוח מחפש מבחינת מיקום. הדירוג מאוד סובייקטיבי. לא יודעים מי דירג, כמה דירגו, ולפי מה. אותה הבעיה נוצרת גם בכל דירוג אחר, לדוגמה דירוג הלקוחות, ניקיון וכו'. גם כאן הדירוג יכול להטעות אם בוצע בצורה שאינה מתאימה לאופי המשתמש ואם כמות המדרגים הייתה קטנה יחסית.

4. ישנן חפיפות- מלונות שמופיעים לפי סוג משתמש אחד וגם לפי סוג משתמש אחר כי הם מתאימים גם וגם – הדבר יכול להטעות מי שמחפש מלון לסוג ספציפי של לקוחות ולא יודע שהמלון שלו מופיע גם במשפחות ולא רק בזוגות לדוגמה .



## דיון כללי

הוויזואליזציה נחוצה מכיוון שאין כיום מענה מספק שנותן את האופציה לראות כמות גדולה של אופציות לינה בהודו על המפה, דבר שחיוני עבור מטיילים ועבור אנשי עסקים הרוצים להשקיע/לבנות/לקנות מלון בהודו. הוויזואליזציה נותנת את המידע הרלוונטי ביותר לשני סוגי משתמשים אלה בהודו – ניקיון המלונות, טיב האזורים, קרבה לנמלי תעופה וכבישים, נופים ותוואי שטח, סוגי הלקוחות שהמלון מתאים להם, גובה השטח, מספר כוכבים ודירוג המלון על ידי מטיילים.

הוויזואליזציה פותרת את המחסור בכלים על ידי מתן המענה לכל המידע הנ"ל בצורה מרוכזת, על מפה ועם אופציה לראות את כל המלונות הללו האחד ליד השני ולבצע השוואה. מהוויזואליזציה ניתן להסיק גם על מגמות כגון הקשר שבין טיב המיקום לדירוג המשתמשים ועוד.

הוויזואליזציה חוסכת זמן רב למשתמש כיוון שלא ניתן להסיק את המידע שניתן להסיק מהוויזואליזציה מקריאת טבלה בעלת 20,000 רשומות. הוויזואליזציה כבר מסננת עבור המשתמש את המידע הרלוונטי לו לפי העדפותיו ולפי אופי המלון שהוא מחפש. בנוסף, למשתמש ניתנת האופציה לבצע תקריב למפה לאזור ספציפי שהוא יודע שבו הוא מעוניין ולראות את כל האופציות שבו, דבר שלא ניתן להסיק מטבלת הנתונים הגולמית.

מבחינת השיפורים להמשך הייתי ממליצה להוסיף מידע על אפשרויות לינה נוספות שאינן מלוונות כיוון שהמידע על מקומות שכאלה הוא דל למרות שהרבה מטיילים מחפשים אופציות זולות יותר כגון בתי הארחה. בנוסף, ניתן להוסיף מידע מספרי לויזואליזציה- ערך הדירוג של המיקום, המלון, הניקיון ועוד. ניתן גם לאסוף מידע על מחירי החדרים ולהוסיף לבסיס הנתונים ולוויזואליזציה- זהו מידע קריטי ביותר עבור מטיילים. רצוי להוסיף לבסיס הנתונים והוויזואליזציה גם קישורים לאתרי הזמנות של המלונות. יש לשקול להחליף את דירוג המיקום של כל מלון בדירוג מיקום כללי ולהציג על המפה צבע לאזור או דירוג מספרי לאזור ואז להשתמש בצבעים כדי להוסיף פרמטר נוסף לוויזואליזציה. אופציה טובה תהייה להוסיף לוויזואליזציה חיפוש מלון לפי שם ספציפי או כתובת.

בסהכ הוויזואליזציה אמינה מאוד ומוסיפה ערך מוסף לכל המחפש מקום לינה בהודו או השקעה במקום שכזה.