ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

2018, 19(3), pp. 585 - 606, DOI: 10.18038/aubtda.398259

GEOSPATIAL SOCIAL NETWORK ANALYSIS WITH USING GIS FOR LOCATION-BASED SERVICE RECOMMENDATION

Levent SABAH [1, *], Mehmet ŞİMŞEK [2]

[1] Department of IT, Düzce University, Düzce, Turkey
[2] Department of Computer Engineering, Faculty of Enginneering, Düzce University, Düzce, Turkey

ABSTRACT

People can mark places they visit and share them with other people with using geospatial online social networks. Discovering socially popular locations on geospatial social networks became more important for many applications such as public transportation system planning, tourist trajectory planning etc. In the literature, there are some studies conducted for this purpose. However, most of these studies focused on only location information and ignored time data of check-in. In this study, we analyzed geospatial social network data by dividing it into time slots. Also, the vast majority of studies in the literature do not provide visual analysis results. This reduces the intelligibility of the results. Our system uses advanced heat maps to provide easy visually interpretable results. The developed system determines the tendency of the check-in intensities at the time of day and the seasons of the year. Using QGIS, which is an open source geographic information system, we obtained check-in data from dataset within Turkey country boundary. Also, Istanbul province's check-in data was used for more specific analyzing with 3-hour ranges of a day. Furthermore, the density of spatial check-in points was obtained by using heat maps.

Keywords: Social network analysis, Spatial data analysis, Geographic ınformation systems, Recommender systems, Heat maps

1. INTRODUCTION

Nowadays, Online Social Networks (OSNs) gives users the opportunity to communicate and share interests with other users. The incorporation of longitude and latitude data triggered new functionalities and introduced the Location-based Social Networks (LBSNs), a kind of OSNs, where users can share geo-tagged information such as check-ins, photos, text etc. [1]. The fusion of mobile devices and social networks is stimulating a wider use of Location Based Services (LBS) and makes it become an important part of our daily life [2]. With the rise of Location-Based Social Networks (LBSNs) which attract lots of new users every day with the potential of bridging the gap between the physical world and digital online social network services [3]. Socially important locations are places that are frequently visited by social media users in their social media life. Discovering socially interesting, popular or important locations from a location-based social network have recently become important for recommender systems, targeted advertisement applications, and urban planning, etc. [4]. Data generated on LBSNs provide rich information on the whereabouts of urban dwellers. Specifically, such data reveal who spends time where, when, and on what type of activity (e.g., shopping at a mall, or dining at a restaurant). That information can, in turn, be used to describe city regions in terms of activity that takes place therein. For example, the data might reveal that citizens visit one region mainly for shopping in the morning, while another for dining in the evening [5].

In this study, we suggest the use of GIS tools for geospatial social network analysis and developed a recommender system. We used real geo-tagged social network dataset to show our approach's abilities. Our system determines the tendency of people to check-in by using foursquare social network data, which is location-based, according to the hours of the day and the months of the year.
Contributions of the paper can be summarized as follows:

- We developed a recommender system with using GIS and spatial social networks together.
- We used time information and geographically labeled data together.

*Corresponding Author: leventsabah@duzce.edu.tr
Received: 23.02.2018 Accepted: 25.06.2018

- We used heat maps for social network analysis.
- We applied the proposed method to a real dataset. We assessed the method for Turkey and İstanbul.

The rest of the paper is organized as follows: in Section 2 we present related work. Section 3 provides material and method. In Section 4, we analyzed the outstanding check-in categories with time-slots. Section 5 provides a detailed analysis of İstanbul. Section 6 summarizes overall results, and Section 7 concludes our paper.

2. RELATED WORK

In this section, we evaluated the related studies in the literature.

Dokuz et al. have developed an algorithm, which uses the gathered dataset from Twitter for identification of mostly visited places in İstanbul [4]. The previous datasets used in the literature were based on GPS data. However, determining the impact of a place needs data such as users' votes, average duration of visits, and density. The authors noted that using this information they obtained more accurate results. In the study, approximately 2500 individuals within the borders of İstanbul were identified with using Twitter API. After that, they have collected tweets of these users. A dataset was created by gathering the contents of these tweets, date/time information and location. Finally, they found important places in İstanbul by analyzing the dataset with the algorithm. However, important places such as Sultan Ahmet Square and Topkapı Palace could not be determined. The authors said that the reason of this is there are not enough tweets among these people about these places.

Cenamor et al. [6] have developed a system called PlanTour, which produces personalized tourist plans with information obtained from the Minube social travel network. Thus, unlike comparable travel suggestion systems, they have obtained travel plans that include important points of the visited city or region. The system identifies the important points of the city or region based on the number of days of travel, and includes roots between points of interest in the tourist plans. The system consists of three sub-services: tourist plan manager, automated planner and viewer. System entries are the city or region, travel dates and possible preferences that the person will visit.

Zheng et al. [7] have aimed that obtaining interesting places and classical travel orders in a specific geographical area with using several people's GPS data in their work. In the study, dataset created from GPS data collected from 107 users for one year is used. In the study, the location histories of several people are modeled with a tree-based hierarchical chart. Based on this model, they have developed HITS (Hypertext Induced Topic Search)-based inference model, which regards a person's access on a location as a directed link from the person to the target.

In their work, Liu and Xiong [8] have developed a system for proposing a prominent point in terms of subject matter and position, using textual and contextual information on the problems of obtaining personalized recommendations. Latent Dirichlet Allocation (LDA) and Topic and Location-aware Probabilistic Matrix Factorization (TL-PMF) methods are used for this. The LDA has been used to learn the interests of users and to identify important points by investigating the textual information associated with these points. TL-PMF has been used to suggest important points. The system uses location-dependent verbal views to avoid the difficulty that textual terms associated with important points are often incomplete or ambiguous.

Chamoso et al. [9] developed a relationship recommender system for the business and employment-focused social network, named as beBee. The system they have developed offers new contacts and jobs to the users by obtaining relevant information from the social network. The recommender system uses job offer disclosures, user profiles, and information gathered from the actions of users. It then applies different criteria to explore possible new ties that may be transformed into relations. In beBee, users can share content and apply for published job offers. In the study, they identified factors that differed from the information provided by the users and analyzed the information contained in the published job offers. One of the difficulties in information extraction is that there is no necessary information for analysis or it is not in a proper structure. For this reason, they have applied text examination and information extraction techniques for the evaluation of users and business connections.

Yu have proposed a new method for networking the geographical area in co-location pattern mining [10]. Co-location patterns represent subsets of logical locational properties that are in spatial proximity [11]. Most of the methods used for co-location mining in the literature have been used for events that occur in a homogeneous and isotopic area with Euclidean distance. It has been stated that the physical movement in the Location Based Services (LBS) is usually restricted by a road network, so that the significance values of co-location patterns involving network-connected events cannot be calculated correctly. The proposed method is based on a network model that uses linear referencing. And it refines neighborhood-related graphics of traditional methods to discover co-location patterns from a large spatial database using network distances instead of Euclidean. In the study, an Edge-based Neighbors Searching (ENS) algorithm has been developed.

Colomo-Palacios et al. [12] they developed a platform named POST-VIA 360 to support the whole life cycle of tourism loyalty after the first visit. Tourists are expressing their views for improving the service quality. They also use these views to influence other tourists' decisions. With this phenomenon, tourists and service providers are generating large amounts of data. POST-VIA 360 was created to help pre-visit, visit and post-visit with a Customer Relationship Management (CRM) approach to improve tourism loyalty and overall performance at every stage of tourists' travel. A case study was conducted, which compares POST-VIA 360 and a group of experts' recommendations to verify the system. And that the suggestions of the system are remarkably accurate compared to previous work done in the field. The most important challenge the system is facing is the size of the dataset it will analyze.

Sattari et al. [13] have proposed a new model to increase location/activity recommendations by integrating additional information from location sources with the dataset obtained from Microsoft Research Asia. Information in the dataset consisting of 2.5 year time interval has been collected from GPS-enabled devices of people visiting Beijing, China. The inputs of the location - activity matrix formed from 167 venues and 5 events (food & beverage, shopping, cinema and demonstration, sports and exercise, tourism and entertainment) consist of the frequency of activities for all people at that location. In addition, the location-property data extracted from the important point database based on the yellow pages of the city has been added to the model to make it more informative. The collected data are expressed in the form of location-feature matrix. Singular Value Decomposition (SVD) was applied to reveal hidden relationships in the data.

Gaete-Villegas et al. [14] have introduced a Location Based Social Network (LBSN) named traMSNet, which implements a matching algorithm considering homophiles as well as the complementary skills of users in the tourist location. They pointed out that existing applications for tourism apply user mapping algorithms from online social networks and that these mapping algorithms are the inherent homophysical concept that similar individuals are more likely to relate to each other. In tourism-focused LBSN applications, the authors have noted that matching in terms of their similarity does not meet the needs of tourists. The main contribution of the work is to suggest a mechanism that matches users, taking into account complementary skills understood from features not shared by different users for a particular location. Results from a survey of 20 volunteers showed that more than two thirds of responders selected complementary skills and a recommendation that combines homophily.

Zheng et al. have developed a personalized friend and location suggestion system for geographic information systems in their work [15]. In this recommendation, firstly, visits made by a particular person to a geographical region in the real world are used as grades in that region. Secondly, the similarity between users is measured by their location history, so that potential group of friends is proposed to each person in a Geographic Information System community. Thirdly, the location of a person in a cluster of unvisited region is estimated by including the location history of that person and the history of other users. It is stated that the Hierarchical-Graph-based Similarity Measurement (HGSM) framework suggests that each individual model their position history uniformly and that the similarity between users can be measured effectively. The proposed system was evaluated by GPS data generated by 75 people at 1 year time interval. As a result, HGSM performs better than similarity measures such as similarity-by count, cosine similarity and Pearson similarity measures.

Based on the literature review, we can summarize the following:

- Often, GPS data is used instead of check-in data. This does not adequately represent the reasons for users to be in that location.
- Generally the datasets used are very small (in some studies, a few dozen of people)
- To the best of our knowledge, there is no study using heat maps for analyzing spatial social networks.

3. MATERIAL AND METHOD

Foursquare is a local search-and-discovery service mobile application, which provides search results for its users. The application provides personalized recommendations for places to go to near a user's current location based on users' "previous browsing history, purchases, or check-in history". Until late July 2014, Foursquare featured a social networking layer that enabled a user to share their location with friends, via the "check-in" - a user would manually tell the application when they were at a particular location using a mobile website, text messaging, or a device-specific application by selecting from a list of venues the application locates nearby. In May 2014, the company launched Swarm, a companion application to Foursquare, which reimagined the social networking and location sharing aspects of the service as a separate application. On August 7, 2014, the company launched Foursquare 8.0, the completely new version of the service which finally removed the check-in and location sharing entirely, to focus entirely on local search (Revolvy web page).

Global-scale check-in dataset includes long-term (about 18 months from April 2012 to September 2013) data collected from Foursquare. It contains 33.278.683 check-ins by 266.909 users on 3.680.126 venues (in 415 cities in 77 countries). Those 415 cities are the most checked 415 cities by Foursquare users in the world. Also, QGIS open source geographic information system is used for spatial analysis in the study.

For social network analysis, we used 6.059.621 check-in data from 2012 and 2013 Foursquare check-in and POI dataset belonging to 35 provinces of Turkey [16, 17]. We especially examined Istanbul province in more detail with 3-hour range of a day.

We can briefly summarize the stages of our study as follows:

- The used dataset in this study is composed of check-ins made worldwide. Firstly, we retrieved the data within the boundaries of Turkey. We used QGIS for this purpose.
- The check-ins in the dataset only have latitude and longitude data. So, the points need to be associated with cities. We conduct an intersection analysis for this purpose.
- After getting spatial check-in points, we divided them into time-slots and generated heat maps for visual social network analysis.

3.1. Spatial Intersection Analysis

It is essential to make the intersection analysis of the points in the dataset to choose the ones inside Turkey's borders and also to determine which points belong to which city. Spatial intersection analysis returns a geometry that represents the shared portion of geometry *A* and geometry *B* [18]. In Figure 1, geometry *A* is a polygon layer, which defines Turkey’s cities and geometry *B* is a point layer, which is a check-in data.

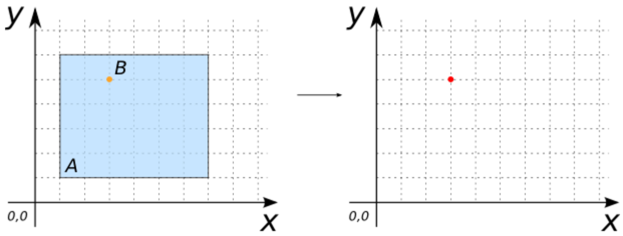


Figure 1. Spatial intersection analysis

Actually, spatial intersection problem is a Point-in-Polygon (PIP) problem of computational geometry. In PIP problem, firstly a polygon (or a hull) is created by using given points. This polygon is boundaries of Turkey (and cities) in our problem. Next, it is calculated whether or not a given coordinate is in the polygon. Also, these coordinates are check-in data in our problem. Let P be a polygon and $q = (x, y)$ a point. We need and external point s, which is $s \neq q$. If we count how much times does $\overline{qs}$ intersect borders of P, we can say that q is an inner point or external point. Let, n be the number of how many times does $\overline{qs}$ intersect borders of P. We can learn the location of the point using (1).

$$PIP(P,q) = \begin{cases} q \text{ is an external point}, & \frac{n}{2} = 0 \\ q \text{ is an inner point}, & \frac{n}{2} \neq 0 \end{cases} \qquad (1)$$

3.2. Creating Heat Maps

We created heat maps to show the intensity of the check-in made according to the regions. For this purpose, we used Kernel Density Method (KDM), which estimates the density of dots falling within a circle with a defined radius and the dot density varying from the center of the circle. Heat map search radius may be in meters or map units. The radius specifies the distance around a point at which the influence of the point will be felt. Larger values result in greater smoothing, but smaller values may show finer details and variation in point density. Default search radius is calculated as in (2).

$$Radius = 0.9 \times min\left(sd, \sqrt{\frac{1}{ln2}} \times md\right) \times n^{-0.2} \qquad (2)$$

Here; sd is standard distance and md is median distance. In KDM, the area is divided into small squares in the form of a grid, and the density is determined by histograms based on the number of dots falling into each square. The distribution frequency of the points is tested by comparing the expected frequency distribution of the squares with the expected value. The density is calculated based on the number of points in a location, with larger numbers of clustered points resulting in larger values. For more details on KDM [19] may be examined. Heat maps allow easy identification of “hotspots” and clustering of [20].

3.3. Dividing Check-ins Into Time Slots

We categorized check-in data according to the time intervals of the day and seasons of the year. Thus, we to determine at which intervals of the day and year the intensive check-ins were made. While heat maps were being created, we created separate layers from all check-in data according to the months of the year and the hours of the day. The heat maps in the Figure 2 show the densities of the check-in points located in the 10 km radius area. Areas marked with red color on the map are identified as the densest areas.

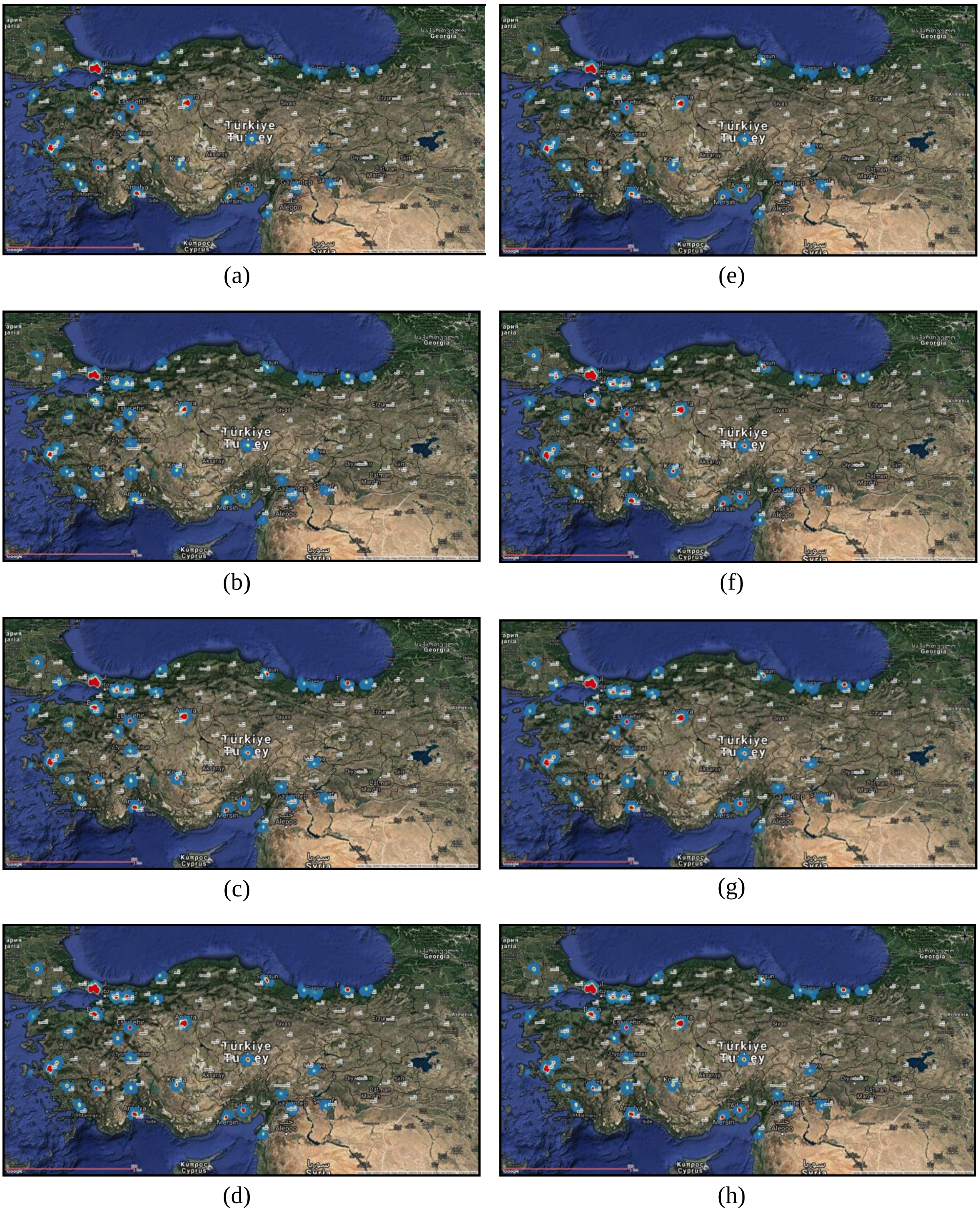


Figure 2. Heat maps by 6-hours and 3-months slot. (a) Heat map of 00:00 – 06:00 time range, (b) Heat map of 06:00 – 12:00 time range, (c) Heat map of 12:00 – 18:00 time range, (d) Heat map of 18:00 – 00:00 time range, (e) Heat map of December, January and February, (f) Heat map of March, April, May, (g) Heat map of June, July, August, (h) Heat map of September, October, December.

We analyzed 4 different time slots: 00:00 – 06:00, 06:00 – 12:00, 12:00 – 18:00, 18:00 – 00:00 and we created separate heat maps with these data. According to this, in the range of 00:00 – 06:00, the maximum number of spot densities in the 10 km area is 2.087. The interval between 06:00 and 12:00 is 14.345 check-in points, which indicates the busiest hour of the day. Furthermore, there are 11.197 check-in points between 12:00 and 18:00 and finally 9,092 check-in points between 18:00 and 00:00. According to this, it is seen that the least intense clock interval of the day is 00:00 – 06:00 and the most intense time range is 06:00 – 12:00. Additionally, we devoted the month of the year into the seasons. Thus, we get 4 seasons: Winter (December, January, February), Spring (March, April, May). Summer (July, August, September), and Autumn (October, November, December). In winter the density of the 10 km area is 10.640 check-in points. In spring, 11.191 check-in points, in summer 4.841 check-in points and in autumn 4.267 check-in points are available. Decrease in check-in intensity during the summer and autumn may be thought because of check-ins at different locations due to the increase in the people's mobility. Likewise, during winter months and spring months, the increase in intensity can be considered as people don't leave much of their locations. Figure 3 shows the heat map, which is created with all the check-ins, the density of the10 km area is 26.234.

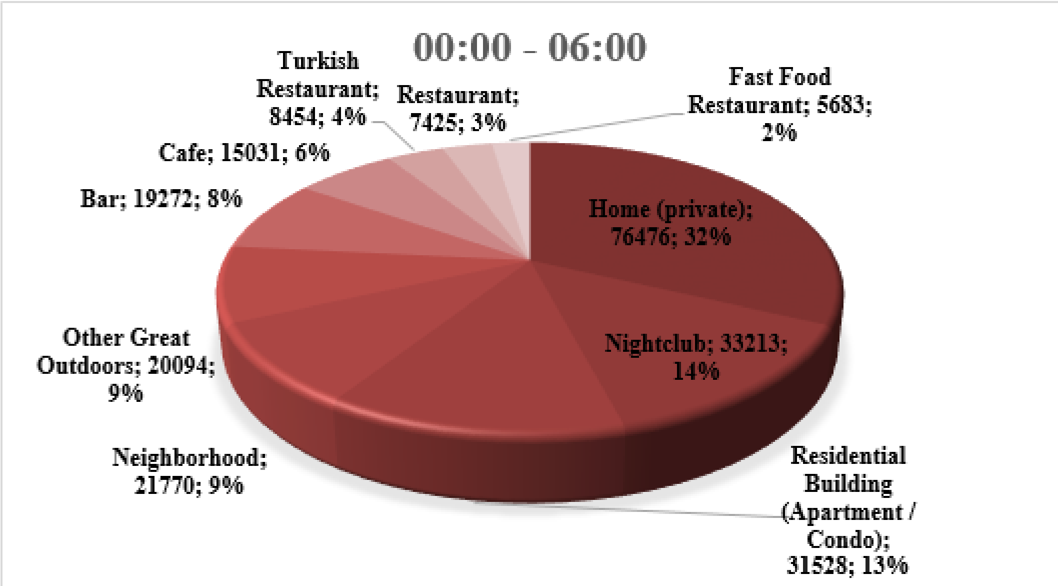
Figure 3. Heat map of all check-in data

4. ANALYSIS OF OUTSTANDING CATEGORIES ACCORDING TO TIME RANGE

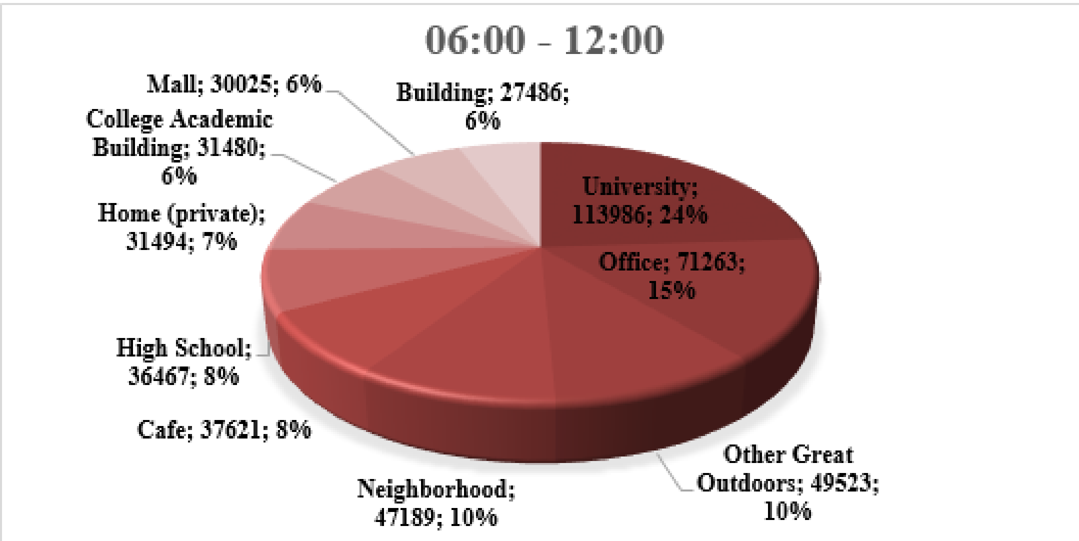
Figure 4a shows check-ins between 00:00 and 06:00. It has been seen that 32% of the first 10 categories with the highest amount are checked-in at home. With the residential building and neighborhood, a total value of 54% is achieved. According to the dataset, more than half of the people are in their homes between 00:00 and 06:00. Besides this, it can be said that the nightclub check-in the second place with 14% and the other bars, cafes and restaurant categories in the other categories together have a moving nightlife with a total value of 37%.

Figure 4b shows check-ins between 06:00 and 12:00. It has been seen that 24% of the top 10 categories are checked-in at the university. With this result, it is seen that the check-in program is generally widespread among the young people with this data obtained in the early hours of the day. With high school and college academic building, 36% of the users are checked-in at the educational institutes of a third of the users. However, 15% of the office check-ins are a significant amount at the overall users.
Figure 4c shows check-ins between 12:00 and 18:00. It has been observed that 19% of the top 10 categories are checked at cafes. Almost all of the check-ins are performed at the same mall. The rate of check-in at mall and cafés is increasing with the decrease in check-in amount made in universities and houses. However, the other great outdoor check-in has increased in the same way. Thus, we can say that people from the school and work went to cafes, shopping centers and outdoors. Figure 4d shows check-ins between 18:00 and 00:00. It has been observed that 21% of the top 10 categories are checked at

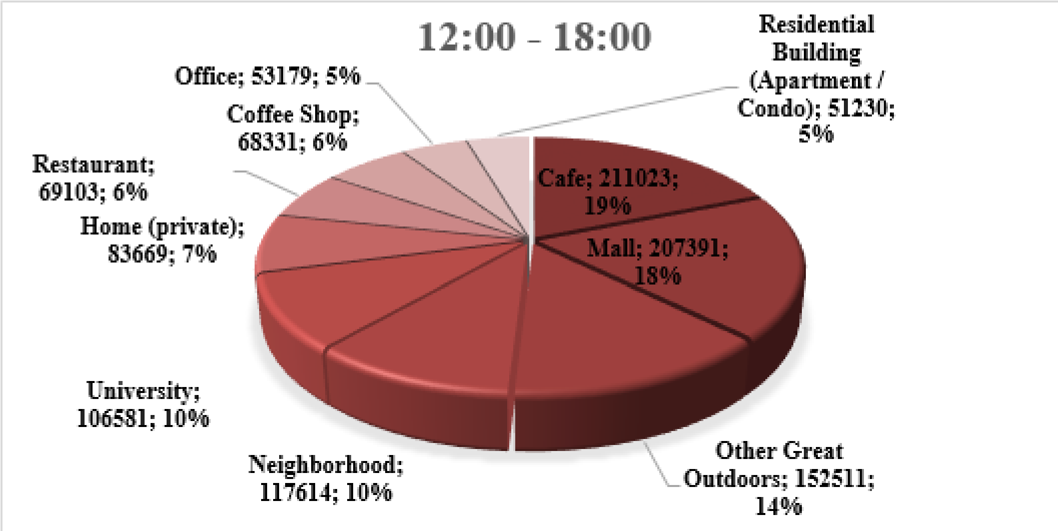
cafés. However, the amount of check-in made at home during this time period is also increasing. Also, it has been seen is that a third of the users are in the houses with a value of 35% together with the neighborhood and the residential building.



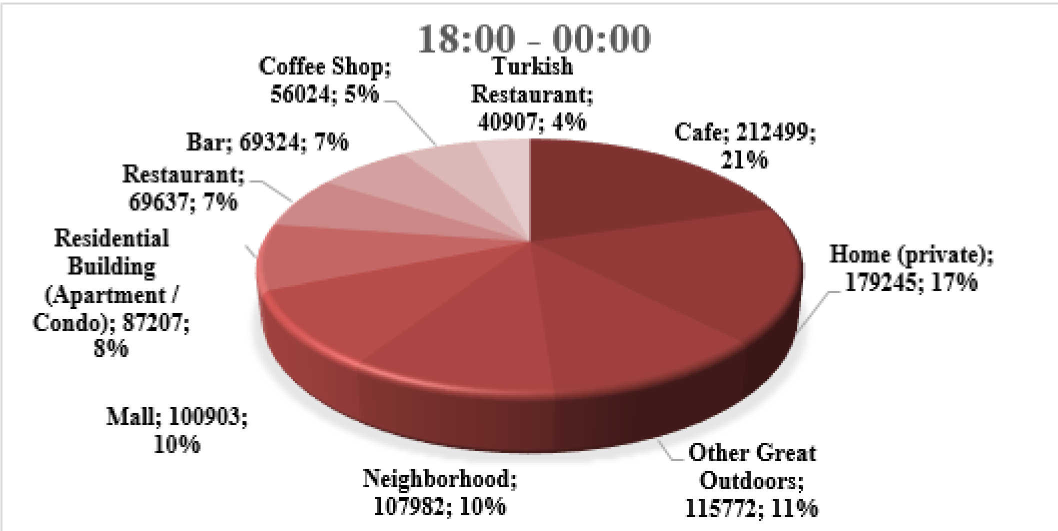
(a)



(b)



(c)



(d)

Figure 4. Check-in category distributions by 6-hour range. (a) 00:00 – 06:00, (b) 06:00 – 12:00, (c) 12:00 – 18:00, (d) 18:00 – 00:00.

5. ANALYSIS OF ISTANBUL PROVINCE

For more specific analysis of check-in dataset, we decide to limit boundaries with Istanbul province. Because İstanbul's check-in data are much more than other provinces. For this analyze, we did spatially intersection process of the dataset with Istanbul province boundaries and check-in dataset points. After analysis process, we get check-in values that are in İstanbul province.

Table 1 shows check-in categories that counts are greater than 27.000. Fewer count categories didn't analyze because simplicity. 9 main categories generated that are mall, work, education, entertainment, accommodation, recreation, sport, transportation and food.

Table 1. Category count ≥ 27.000

Category	Sub Categories	Count
Mall	Mall	339.409
Work	Office	136.284
	Plaza	62.273
	Co-working Space	45.208
	Factory	35.727
Education	University	233.669
	College Academic Building	63.954
	Student Center	53.331
	High School	51.167
	College Cafeteria	30.165
Entertainment	Bar	111.398
	Nightclub	53.273
	Pub	39.406
	Movie Theater	27.581
Accommodation	Home (private)	370.884
	Residential Building (Apartment / Condo)	191.492
	Building	82.384
	Hotel	45.600
	Housing Development	40.501
Recreation	Other Great Outdoors	337.900
	Park	49.904
Sport	Gym	48.273
	Gym / Fitness Center	39.280
	Soccer Stadium	31.424
Transportation	Road	94.487
	Bus Station	72.930
	Airport	45.832
Food	Cafe	476.174
	Restaurant	156.874
	Coffee Shop	137.781
	Turkish Restaurant	104.132
	Fast Food Restaurant	81.570
	Dessert Shop	46.767
	Breakfast Spot	31.873
	Bakery	29.550

Figure 5 shows 9 categories with a percentage. Top 3 counted check-in values are food (28%), accommodation (19%) and education (12%) respectively.

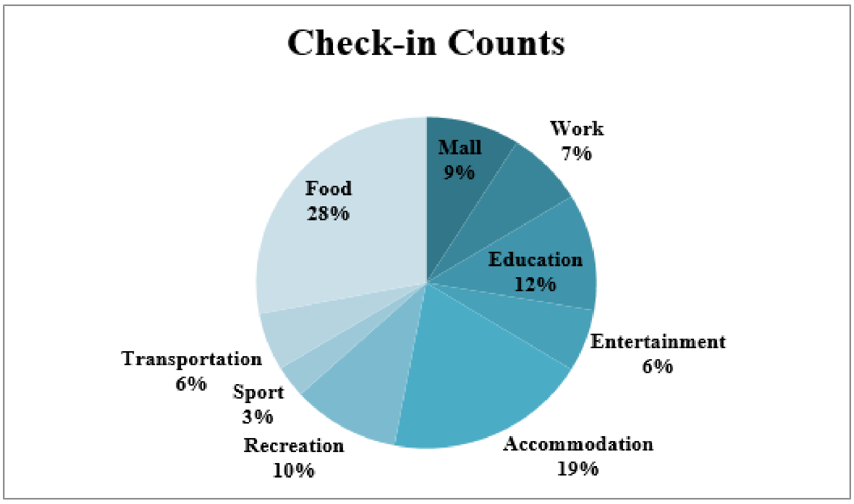


Figure 5. Top 9 check-in categories with percentage

5.1. Analysis of İstanbul Province Check-In Categories According to Time Range

For Istanbul province analysis, 24 hours of the day is divided into 8 time-slots. 9 categories, which are analyzed with percentages, are shown in Table 1.

Figure 6a shows the percentage of check-in data between 00:00 and 03:00. It is seen that entertainment is the highest category with 42%. Entertainment, food, and recreation categories have a total share of 76%. As a result, a lively nightlife can be mentioned.

Figure 6b shows the percentage of check-in data between 03:00 and 06:00. Again entertainment is the highest category with 48%. Entertainment, food, and recreation categories have a total of 73% of the total; it is observed that this time period also continued to be active. Figure 6c shows the percentage of check-in data between 06:00 and 09:00. It is seen that education is the highest category with 29%. Nightlife activity at 00:00 to 06:00 hours falls 22% in this hour range. However, the percentage of education and work is increasing with 52%.

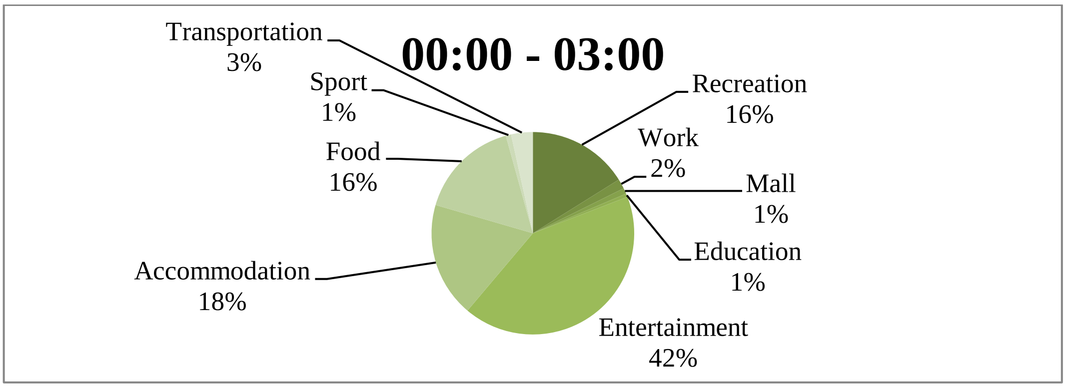
Figure 6d shows the percentage of check-in data between 09:00 and 12:00. It is seen that education is the highest category with 35%. Education and work rate is 55%. In addition, mall check-ins are increasing with 9%.

Figure 6e shows the percentage of check-in data between 12:00 and 15:00. It is seen that mall is the highest category with 24%. It is seen that the rate of shopping, food, and recreation is increasing while the working rate is decreasing in the time interval which corresponds to the lunchtime.

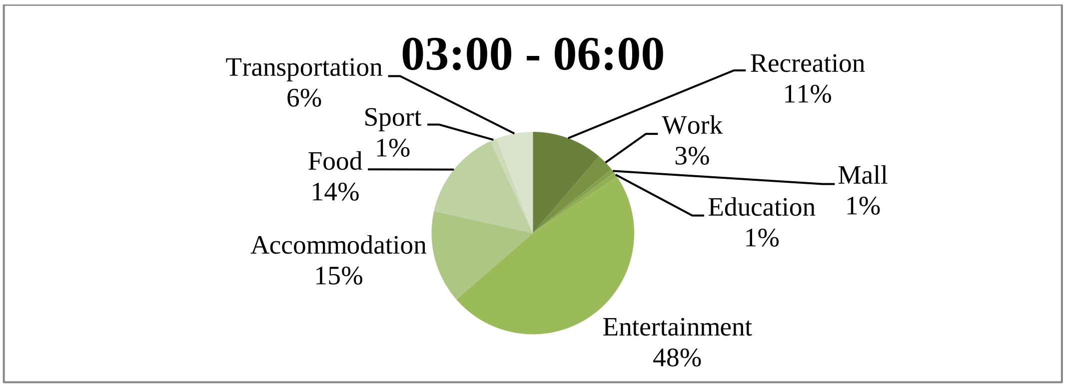
Figure 6f shows the percentage of check-in data between 15:00 and 18:00. It is seen that food and mall are the highest categories with 25%. It is seen that afternoon mainly shopping, food and recreation categories are increasing by 69%.

Figure 6g shows the percentage of check-in data between 18:00 and 21:00. It is seen that food is the highest category with 28%. After 18:00, shopping, food and recreation increased also entertainment and sports categories increased. It is thought that the decrease in the work category is the effect of the employees in the increase of the entertainment and sports categories.

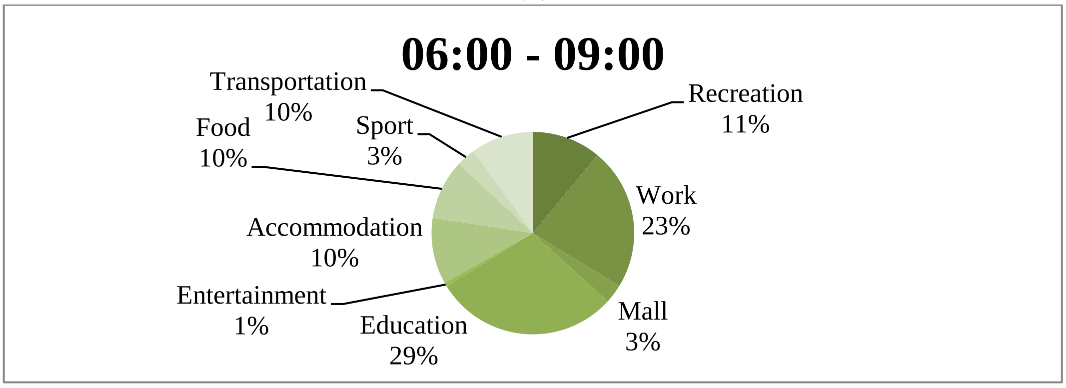
Figure 6h shows the percentage of check-in data between 21:00 and 00:00. It is seen that food is the highest category with 24%. It seems that check-ins in the mall category have decreased as the shopping centers start to close after 21:00. Food categories are most intense categories. Entertainment and accommodation categories are also increasing in similar proportions. This can be interpreted as touristic and business people coming to Istanbul at similar rates.



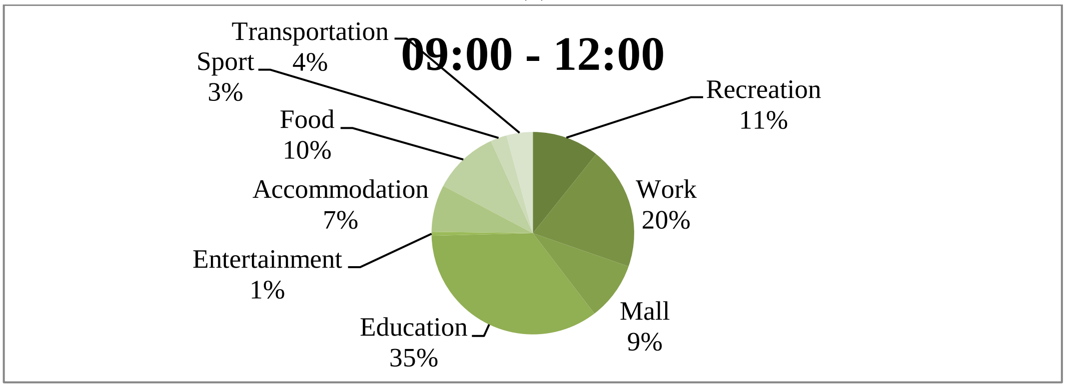
(a)


03:00 - 06:00
Transportation 6%
Sport 1%
Food 14%
Accommodation 15%
Recreation 11%
Work 3%
Mall 1%
Education 1%
Entertainment 48%

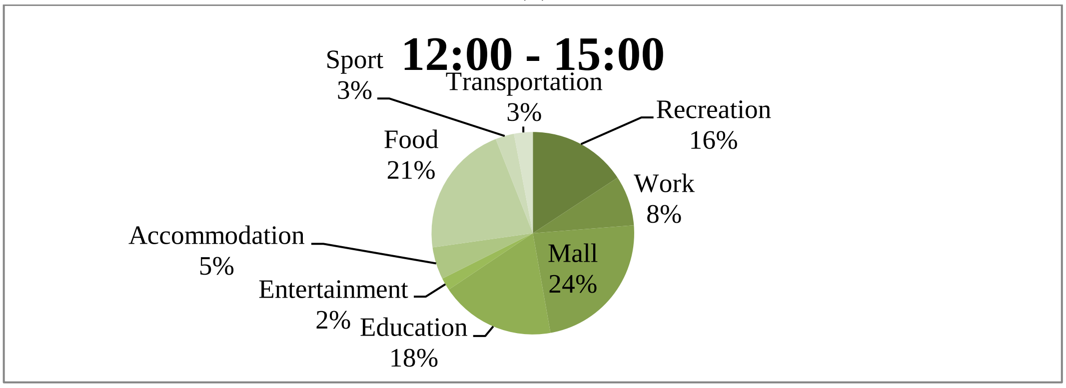
(b)


06:00 - 09:00
Transportation 10%
Food 10%
Sport 3%
Accommodation 10%
Entertainment 1%
Education 29%
Recreation 11%
Work 23%
Mall 3%

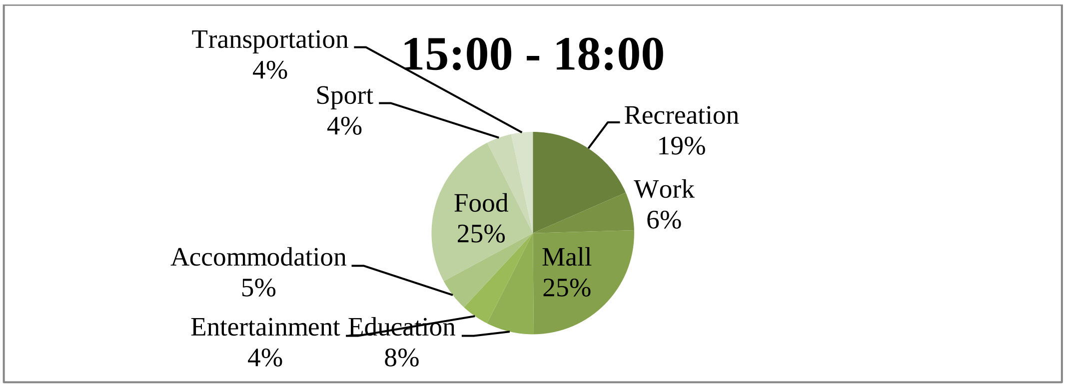
(c)


09:00 - 12:00
Transportation 4%
Sport 3%
Food 10%
Accommodation 7%
Entertainment 1%
Education 35%
Recreation 11%
Work 20%
Mall 9%

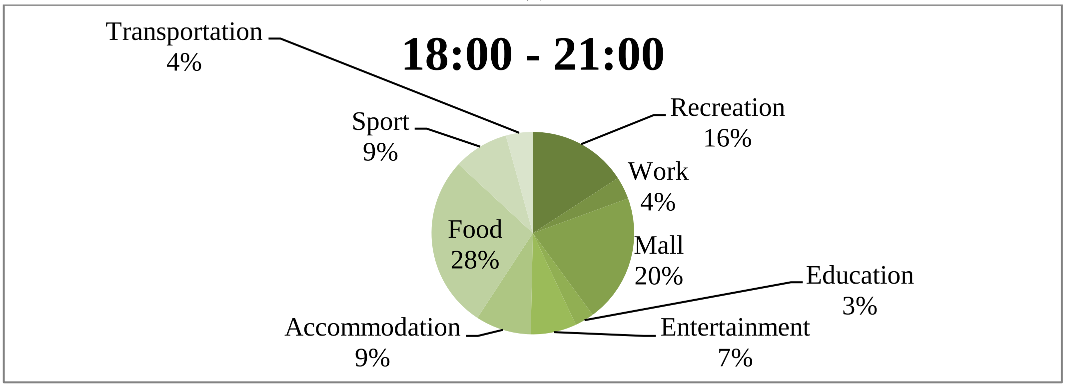
(d)


12:00 - 15:00
Sport 3%
Transportation 3%
Food 21%
Accommodation 5%
Entertainment 2%
Education 18%
Recreation 16%
Work 8%
Mall 24%

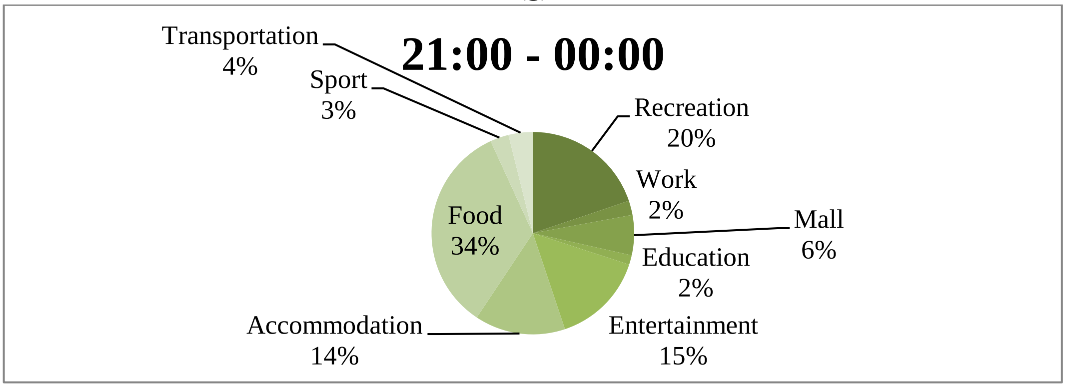
(e)



(f)



(g)



(h)

Figure 6. Check-in category distributions by 3-hour range. (a) 00:00 – 03:00, (b) 03:00 – 06:00, (c) 06:00 – 09:00, (d) 09:00 – 12:00, (e) 12:00 – 15:00, (f) 15:00 – 18:00, (g) 18:00 – 21:00, (h) 21:00 – 00:00.

When the heat map analysis is done, the categories with the number of check-in points of 27.000 and above are considered. These categories are combined with top categories. Table 1 shows these 9 categories that are Mall, Work, Education, Entertainment, Accommodation, Recreation, Sport, Transportation, Food. Table 2 shows that generated heat maps' maximum density of each category with hour ranges that 00:00 – 03:00, 03:00 - 06:00, 06:00 - 09:00, 09:00 - 12:00, 12:00 - 15:00, 15:00 - 18:00, 18:00 - 21:00 and 21:00 - 00:00.

Table 2. Maximum heat map density point of categories hour ranges

	00:00 - 03-00	**03:00 - 06:00**	**06:00 - 09:00**	**09:00 - 12:00**	**12:00 - 15:00**	**15:00 - 18:00**	**18:00 - 21:00**	**21:00 - 00:00**
Recreation	639,993	123,017	287,832	999,995	2229,69	2338,69	1921,76	1560,53
Work	62,397	30,6887	594,792	1823,41	1132,26	775,085	434,925	180,591
Mall	36,4808	7,32304	79,1262	865,287	3321,63	3209,07	2489,11	504,951
Education	21,7174	8,66061	769,747	3256,99	2601,46	964,437	374,295	118,243
Entertainment	1668,66	521,725	16,6323	58,4485	284,508	561,531	886,543	1169,36
Accommodation	728,557	161,37	276,538	700,702	730,779	663,503	1077,91	1139,35
Food	643,165	159,007	251,467	972,847	3011,19	3225,65	3358,3	2663,43
Sport	33,5761	11,57	78,31	243,766	421,094	507,423	1068,62	233,501
Transportation	137,447	63,3084	261,202	390,41	424,558	435,829	521,81	305,241

Figure 7 shows the legend of generated heat maps have been colored towards the red tones from the blue tones. The densities are made according to the point amounts within 2 km radius. Zone intensities are increasing towards blue to red. Zone intensities are divided to 5 ranges. Ranges were created 0, 125, 250, 375 and 500 respectively.

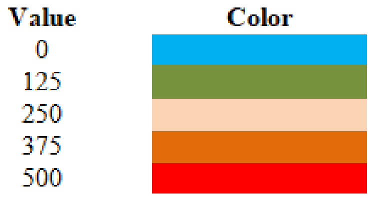


Figure 7. Legend of heat maps

5.2. Heat Maps by Check-in Categories

5.2.1. Accommodations

Sub-categories of check-ins in accommodation category are home (private), residential building (apartment/condo), building, hotel and housing development. The heat maps in Figure 8 show that the busiest hours people spend at home are between 18:00 and 00:00. The number of check-ins between 03:00 and 06:00 is very low. This situation may be due to people's sleep. However, there may be people who spend time elsewhere during these hours. To understand this, other check-in categories for the same hours need to be examined.

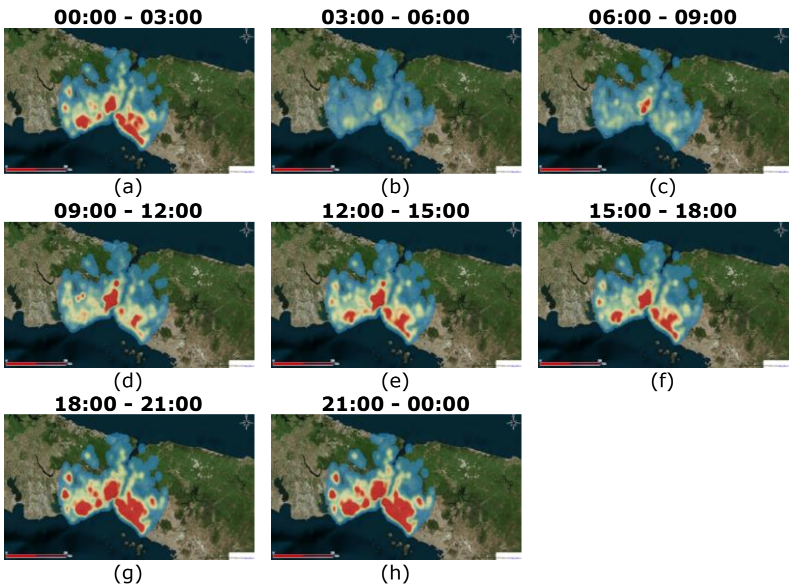


Figure 8. Heat maps of accommodation category check-ins

5.2.2. Recreation

Sub-categories of check-ins in recreation category are other great outdoors and park. According to the heat maps in Figure 9, recreational activities seem to start at noon and intensify in the evening hours. It can be thought that most of the time when people are studying or working in the day, they spend time at the end of the day.

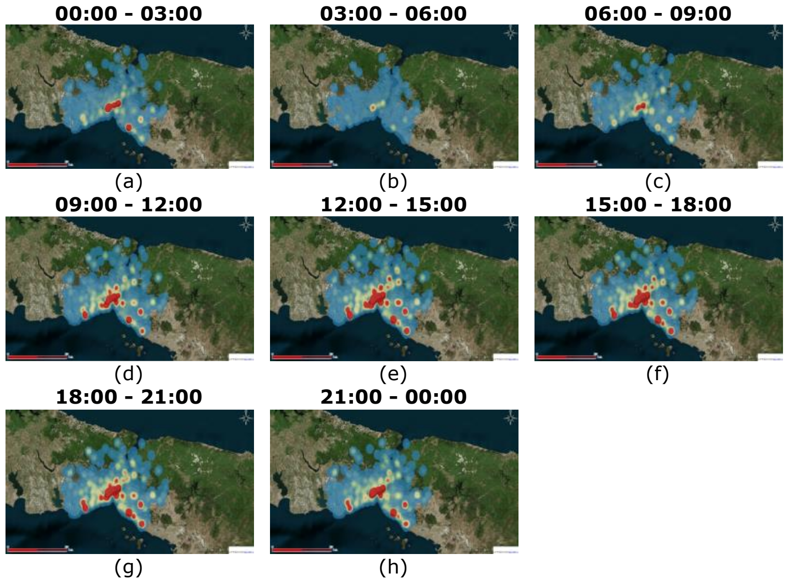


Figure 9. Heat maps of recreation category check-ins

5.2.3. Work

Sub-categories of check-ins in work category are office, plaza, co-working space and factory. According to the heat maps in Figure 10, the working life generally appears to be concentrated between 09:00 and 18:00. Apart from that, there is also a significant concentration between 06:00 - 09:00. It can be thought that this is mainly due to shift-workers and service workers.

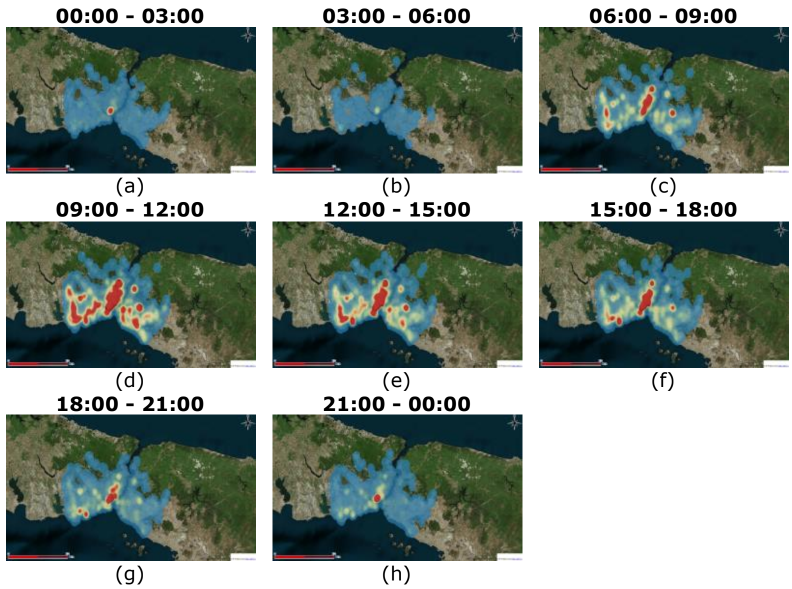


Figure 10. Heat maps of work category check-ins

5.2.4. Mall

Mall category was evaluated solely as a category. According to the heat maps in Figure 11, people are shopping most intensively between 12:00 and 21:00. There is a decrease in density after 21:00. This maybe because the shopping centers are usually opens until 22:00.

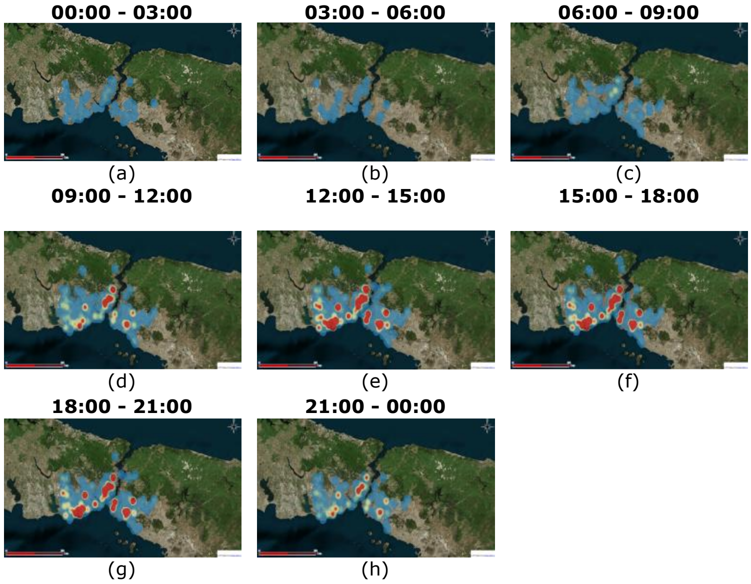


Figure 11. Heat maps of mall category check-ins

5.2.5. Education

Sub-categories of check-ins in education category are university, college academic building, student center, high school, and college cafeteria. According to the heat maps in Figure 12, it is seen that the training activities are concentrated in the range of 06:00 to 18:00. It can be considered that the reason for the relatively low interval between 06:00 - 09:00 is usually only the early training hours of primary school high school students.

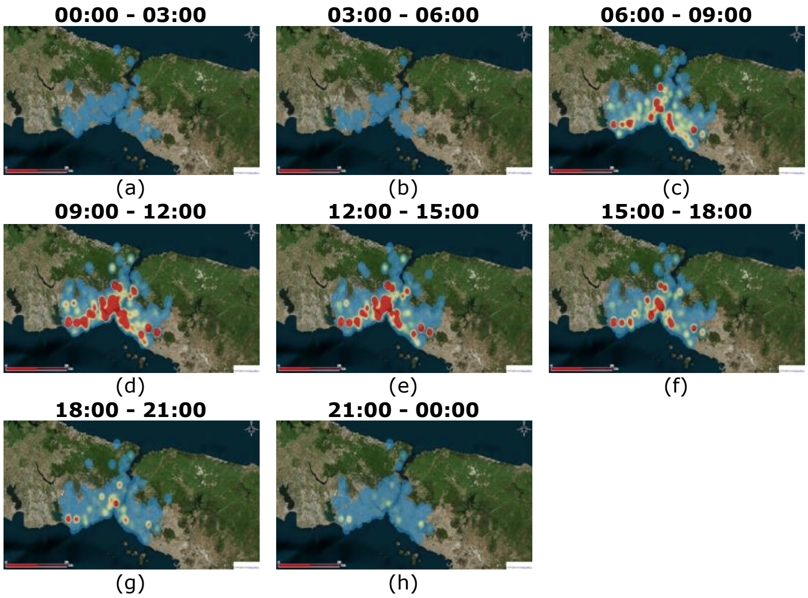


Figure 12. Heat maps of education category check-ins

5.2.6. Entertainment

Sub-categories of check-ins in entertainment category are bar, nightclub, pub, movie theater. According to the heat maps in Figure 13, in the entertainment category, check-ins are generally seen to be concentrated outside the hours of 06:00 to 12:00 hours. It is thought that there is a density spreading throughout the day with the reason that there are entertainment places that can be visited almost every hour of the day.

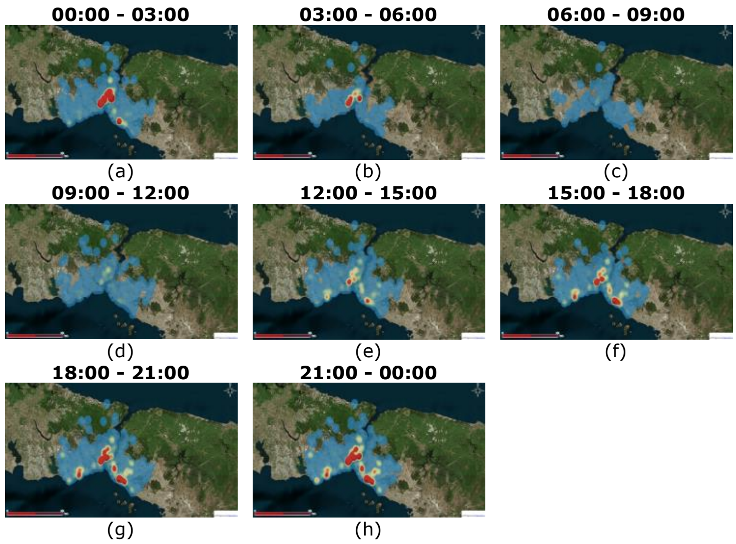


Figure 13. Heat maps of entertainment category check-ins

5.2.7. Food

Sub-categories of check-ins in food category are café, restaurant, coffee shop, Turkish restaurant, fast food restaurant, dessert shop, breakfast spot and bakery. According to the heat maps in Figure 14, it is observed that the check-in density of the food category is high during the day except for the period between 03:00 - 09:00 hours. It is thought that the density between 12:00 - 00:00 starts with lunch break. Out of work and training hours, the intensity of food & beverage places such as café, restaurant has increased.

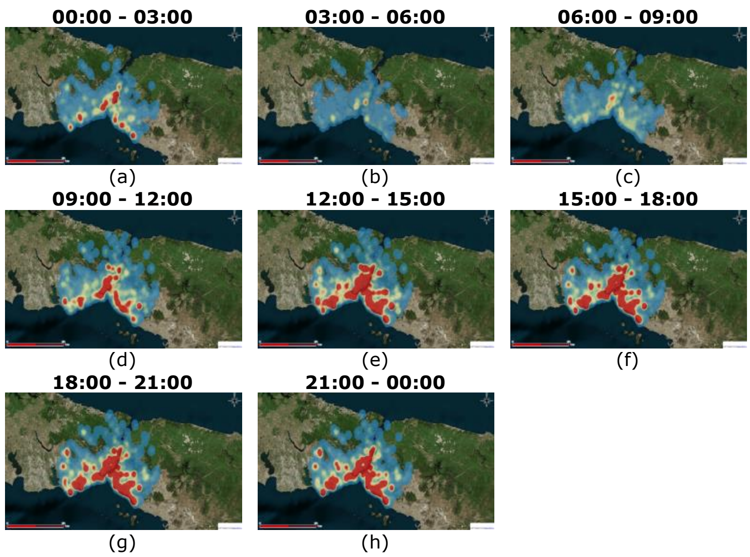


Figure 14. Heat maps of food category check-ins

5.2.8. Sport

Sub-categories of check-ins in sports are gym, gym/fitness center and soccer stadium. According to the heat maps in Figure 15, it can be considered that sports activities concentrate between the opening and closing times of sports halls. Football matches are usually played in the evening. It can be said that the density between 18.00 and 21.00 is due to this. In addition, check-ins are concentrated at the locations of the major soccer stadiums in Istanbul. This supports our hypothesis.

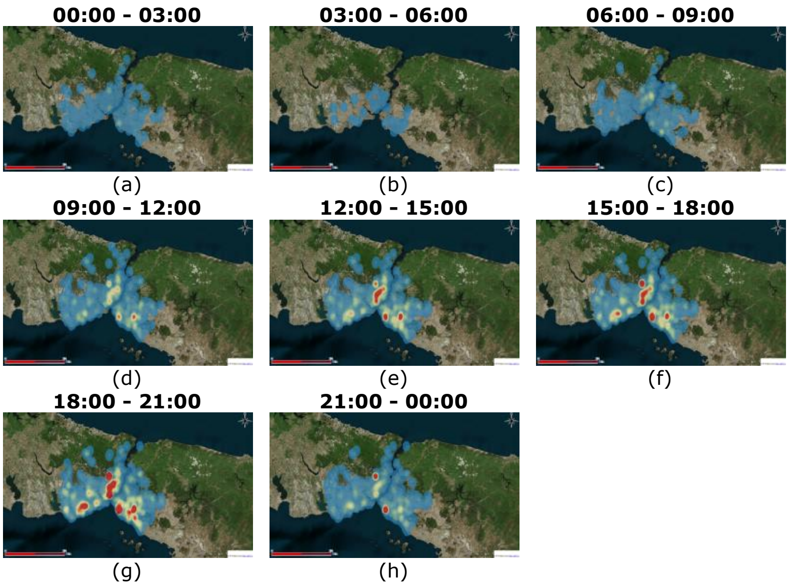


Figure 15. Heat maps of sport category check-ins

5.2.9. Transportation

Sub-categories of check-ins in transportation category are road, bus station and airport. According to the heat maps in Figure 16, it appears that the check-ins in the transportation category are concentrated outside the range 00:00 to 06:00. The reason for this may be that public transport services are not offered at these times and there is no demand for public transport services.

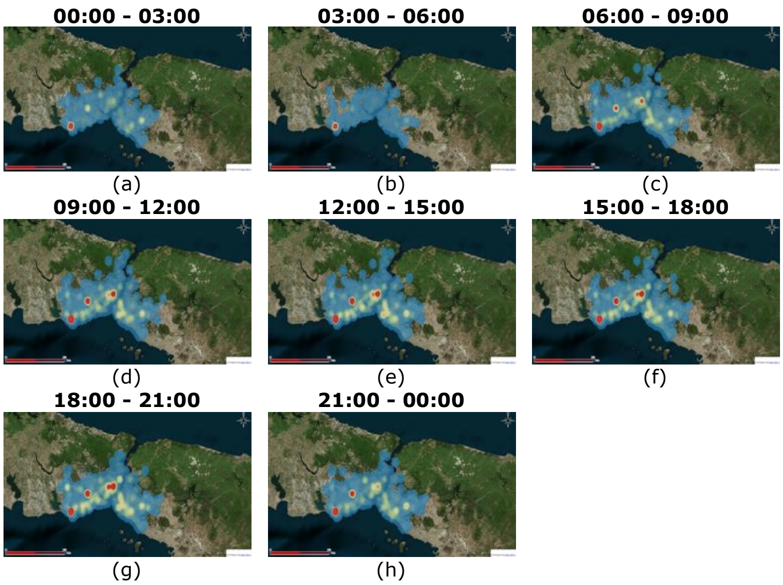


Figure 16. Heat maps of transportation category check-ins

6. EXPERIMENTAL RESULTS AND DISCUSSION

The results as a whole are evaluated in Table 3. First 10 most checked-in categories are café, coffee shop, home (private), mall, neighborhood, office, other great outdoors, residential building (apartment/condo), restaurant and university. The total number of check-ins done on provincial basis shows that three major cities of Turkey are in the foreground. İstanbul is leader on check-ins. It is followed by Ankara and İzmir. Figure 17 shows the percentage of the total number of check-ins made in the first 10 categories which are checked most in number İstanbul has 53% of the check-ins made. It is followed by İzmir with 11 percent and Ankara with 9 percent respectively. Figure 18 shows the distribution of check-ins in provinces based on top 10 categories which are mostly checked-in. The horizontal label refers cities of Turkey in Figure 18.

In the analysis of İstanbul, it has been seen that entertainment category has the highest rate in the first quarter of the day. It is also has been seen that by adding food and recreation categories, it is a lively nightlife with a rate of over 70%.

In the second quarter of the day, working and education category ratios have been seen to increase with half of all check-ins. In addition, shopping rates at noontime are also increased.

In the third quarter of the day, food and mall check-ins constitute half of all check-ins. There is also a decrease in the number of work check-ins.

In the last quarter, entertainment category increased by 4% to 15%. The increase in the food category, which has increased afternoon, also continues. The increase in the category of accommodation can be interpreted as the attraction center of İstanbul in terms of tourism and business. There are no significant changes in other categories.

Table 3. The number of check-ins made in mostly checked-in 10 categories.

City Code	Cafe	Coffee Shop	Home (private)	Mall	Neighborhood	Office	Other Great Outdoors	Residential Building (Apartment / Condo)	Restaurant	University	Total
1	14064	5741	11943	4504	2020	2654	3613	5946	1560	3453	55498
3	2080	345	777	966	216	275	82	404	138	1032	6315
6	40482	17130	33822	41065	19923	10182	19033	20394	11682	17878	231591
7	8126	4259	14075	11100	5600	3832	2386	7079	4045	1678	62180
9	4375	265	2261	1854	295	143	2427	1215	492	711	14038
10	9283	421	3358	1596	912	271	1371	1473	2064	1596	22345
14	2433	119	965	840	85	144	219	383	235	577	6000
15	1		2		1				17		21
16	18739	4957	11637	14809	11949	4055	3637	8289	5861	4891	88824
17	243	15	995	363	105	114	430	469	104	863	3701
20	8328	894	6086	4851	2043	721	2319	3481	970	1750	31443
22	4677	339	2298	1353	93	188	2286	1465	352	934	13985
26	8445	2861	7677	6010	846	1247	5277	3143	4952	3691	44149
27	6535	488	3086	2062	1257	1078	611	2559	1343	1903	20922
28	346		923	196	322	45	225	464	75	32	2628
31	1308	175	1064	438	226	126	197	731	145	204	4614
32	2915	213	1855	934	539	37	894	1274	231	258	9150
33	7890	2296	5799	8883	594	1608	2286	7116	1961	1694	40127
34	197551	71016	166639	171150	182528	94508	224901	77531	88303	153715	1427842
35	55932	11153	43200	29838	49680	8123	38573	17262	15503	15152	284416
38	4241	2383	3721	3746	1132	1022	520	2686	1030	1849	22330
41	12372	2333	7577	6591	3337	1271	5404	5710	3104	3609	51308

City Code	Cafe	Coffee Shop	Home (private)	Mall	Neighborhood	Office	Other Great Outdoors	Residential Building (Apartment / Condo)	Restaurant	University	Total
42	5266	2388	3605	3343	1159	1069	817	3491	1126	1409	23673
43	3591	260	2232	1047	113	193	363	1102	199	1402	10502
44	1377	551	655	698	220	116	965	576	145	304	5607
45	4532	355	2729	1419	237	176	2604	1047	321	1516	14936
46	1094	147	686	182	142	150	93	1059	69	274	3896
48	5335	245	1649	267	99	51	1313	1077	591	1282	11909
52	1054	295	807	113	214	193	291	435	404	113	3919
53	1809	8	1173	623	547	214	596	656	197	112	5935
54	13017	1597	12512	4371	1212	837	4267	3587	4421	4268	50089
55	8580	1205	4247	4685	1444	672	2863	3458	1443	1531	30128
59	4005	217	2340	2017	1561	157	2607	1206	456	715	15281
61	12892	2766	6936	6384	3398	623	3709	3946	3007	2974	46635
63	1405	2	1065	495	313	136	287	436	138	70	4347
67	1851	342	486	616	193	53	434	342	189	229	4735
81			2								2
Total	**476174**	**137781**	**370884**	**339409**	**294555**	**136284**	**337900**	**191492**	**156873**	**233669**	**2675021**

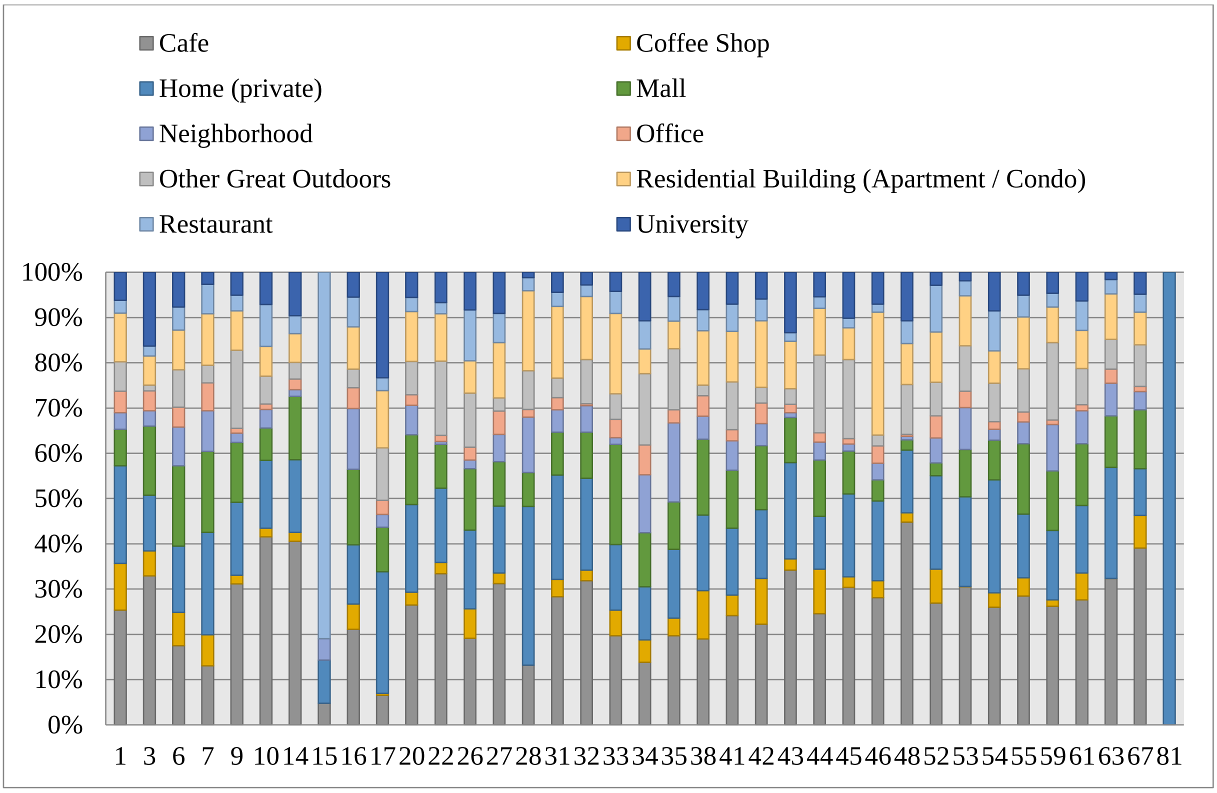


Figure 17. The distribution of check-ins on mostly checked-in 10 categories in provinces.

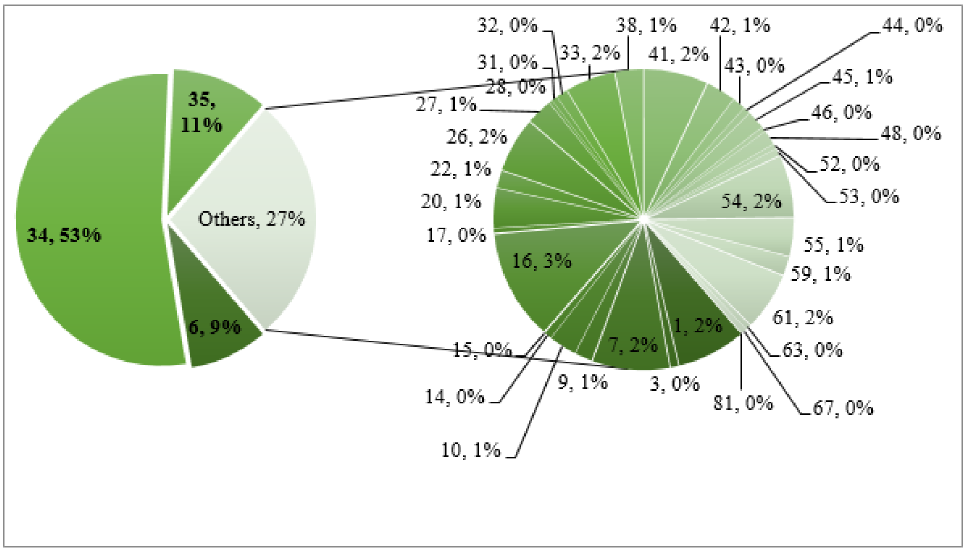


Figure 18. The percentage of the total number of check-ins made in mostly checked-in 10 categories.

7. CONCLUSION

In this study, we developed a system, which uses GIS tools and abilities for social network analysis; we analyzed geospatial social network data by dividing it into time slots. Also, we used advanced heat maps to produce easy interpretable visual results. Developed system can give benefits in many different fields. For example, travel (or tourist) trajectories, activities according to seasons and time of the day, breakpoints, accommodation etc. can be planned by this system. Public authorities can use this system to improve the quality of public transport services. Another possible usage of the system is investment planning. With using this system, user check-ins can be analyzed to determine the best locations for an investment. Thus, it is possible to make the most suitable place selection by evaluating the regions that the investors will address. As a result, both a profitable business opportunity can be caught and an unnecessary investment cost can be reduced.

An advantage of the system we have developed is that it does not require high processing power. We analyzed more than 6 million check-in data in an acceptable time using a standard desktop computer. As a future work, we are planning to develop an online tool with using GIS abilities and advanced heat maps for spatial social network analysis.

REFERENCES

[1] Kefalas P, Manolopoulos Y. A time-aware spatio-textual recommender system. Expert Systems With Applications 2017; 78: 396–406.

[2] Sun Y, Chen M, Hu L, Qian Y, Mehedi M. ASA : Against statistical attacks for privacy-aware users in Location Based Service. Future Generation Computer Systems 2017;70: 48–58.

[3] Assem H, Xu L, Buda T S, O’Sullivan D. Spatio-Temporal Clustering Approach for Detecting Functional Regions in Cities. In: 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI); 2016, pp. 370–377.

[4] Sakir A, Celik M. Discovering socially important locations of social media users. Expert Systems With Applications 2017; 86 (December 2015):113–124.

[5] Celikten E, Le Falher G, Mathioudakis M.Modeling Urban Behavior by Mining Geotagged Social Data. IEEE Transactions on Big Data 2017; 3(2):220–233.

[6] Cenamor I, de la Rosa T, Núñez S, Borrajo D. Planning for tourism routes using social networks. Expert Systems with Applications 2017; 69: 1–9.

[7] Zheng Y, Zhang L, Xie X, Ma W-Y. Mining interesting locations and travel sequences from GPS trajectories. Proceedings of the 18th international conference on World wide web 2009; 49:791.

[8] Liu B, Xiong H. Point-of-Interest Recommendation in Location Based Social Networks with Topic and Location Awareness. Sdm 2013; 396–404.

[9] Chamoso P, Rivas A, Rodríguez S, Bajo J. Relationship recommender system in a business and employment-oriented social network. Information Sciences 2018; 433–434: 204–220.

[10] Yu W. Spatial co-location pattern mining for location-based services in road networks. Expert Systems with Applications 2016; 46:324–335.

[11] Huang Y, Shekhar S, Xiong H. Discovering colocation patterns from spatial data sets: a general approach. IEEE Transactions on Knowledge and Data Engineering 2004;16(12): 1472–1485.

[12] Colomo-Palacios R, García-Peñalvo F J, Stantchev V, Misra S. Towards a social and context-aware mobile recommendation system for tourism. Pervasive and Mobile Computing 2017; 38: 505–515.

[13] Sattari M, Manguoglu M, Toroslu I H, Symeonidis P, Senkul P, Manolopoulos Y. Geo-activity recommendations by using improved feature combination. Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12) 2012; 996–1003.

[14] Gaete-Villegas J, Cha M, Lee D, Ko I.-Y. TraMSNET: A Mobile Social Network Application for Tourism. Proceedings of the 2012 ACM Conference on Ubiquitous Computing 2012; 1004–1011.

[15] Zheng Y, Zhang L, Ma Z, Xie X, Ma W.-Y. Recommending friends and locations based on individual location history. ACM Transactions on the Web 2011; 5(1):1–44.

[16] Yang D, Zhang D, Qu B. Participatory Cultural Mapping Based on Collective Behavior Data in Location Based Social Networks. ACM Trans. on Intelligent Systems and Technology (TIST) 2015.

[17] Yang D, Zhang D, Chen L, Qu B. Nation Telescope: Monitoring and Visualizing Large-Scale Collective Behavior in LBSNs. Journal of Network and Computer Applications (JNCA) 2015; 55:170-180.

[18] Yeung A K W, Hall G B. Spatial Database Systems. Springer, 2007.

[19] Silverman B W. Density Estimation for Statistics and Data Analysis. London, UK: Chapman and Hall, 1986.

[20] Menke K, Smith R, Pirelli L, Hoesen J V. Mastering QGIS – Second Edition. Packt, 2016.