

Analysis of Crimes against Male-Female Employment and Tourism in SriLanka

1st Waruna Priyankara

dept. Computer Science and Engineering
University of Moratuwa
Moratuwa, SriLanka
199371p@uom.lk

3rd Pumudinee Kumarasiri

dept. Computer Science and Engineering
University of Moratuwa
Moratuwa, SriLanka
199338x@uom.lk

2nd Prageeth Anjula

dept. Computer Science and Engineering
University of Moratuwa
Moratuwa, SriLanka
199304p@uom.lk

4th Hesitha Wijayasinghe

dept. Computer Science and Engineering
University of Moratuwa
Moratuwa, SriLanka
199372u@uom.lk

Abstract—A significant rise in crime-related incidents is always a major problem for any country. Specially, a developing country like Sri Lanka suffers a lot from crimes. This affects Sri Lanka's economic growth and has a reasonable impact on continuously growing social problems. Theoretical and application-oriented approaches which provide insights into why and where crimes take place are more valuable for decision and policy makers. Geographic information systems and graphical analysis are proven to be essential for studying criminal activities. This paper explores how criminal data relate to male-female employment and tourism industry in Sri Lanka.

Index Terms—data science, hypothesis, analysis, statistics

I. INTRODUCTION

Number of data sets are freely available at Open Data Portal [2] of SriLanka which is an initiative of SriLanka government to facilitate researches, policy makers and decision makers to come up with ideal solutions for prevailing problems. Crime is a serious problem in SriLanka. Crime related news are published in newspapers each and every day. Authors have chosen data mentioned in Table “Tab. I” to analyze about criminal activities and their relation between tourism and employment. Before giving solutions for criminal activities the true reasons behind crimes should be understood. Authors want to see whether there are correlation of crime data with employment and tourism. These data has been recorded as district base. Before analyzing three sets were integrated together. The primary contributions of this paper are:

- Analyze crime data and male-female employment data for correlation
- Analyze crime data and tourism data for correlation
- Use scientific methods in data science for processing, analyzing and visualization

TABLE I
DATA SET DETAILS

| Data Set | Link |
|--|---|
| Crime data 2010-2012 | http://www.data.gov.lk/dataset/crime-data-2010-2012 |
| Employment by Industrial Sector 2012 | http://www.data.gov.lk/dataset/employees-industrial-sector-2012 |
| Accommodation information for tourists | http://www.data.gov.lk/dataset/accommodation-information-tourists |
| Population by district in census years | http://www.data.gov.lk/dataset/population-district-census-years |

II. DATA AND METHODOLOGY

A. Data

Data set and reference to open data portal is mentioned in Table “Tab. I”. Data set dimensions are mentioned below

- Crime Data set size = (25, 23)
- Employment Data set size = (25, 13)
- Tourism Data set size = (2130, 7)
- Integrated dataframe shape = (25, 36)

Robbery, abduction, rape crime data distribution throughout the country were plotted. How male-female employment vary across district and room count distribution were also plotted “Fig. 1”. We can observe that the employed female population is higher in district where abduction/kidnapping is also high. At the same time hotel room count is also higher in district where robbery and other crimes are high. We assume tourism activities are high in districts where totally offered hotel room count is high and it is taken as a measure of the scale of the tourism activity and attraction.

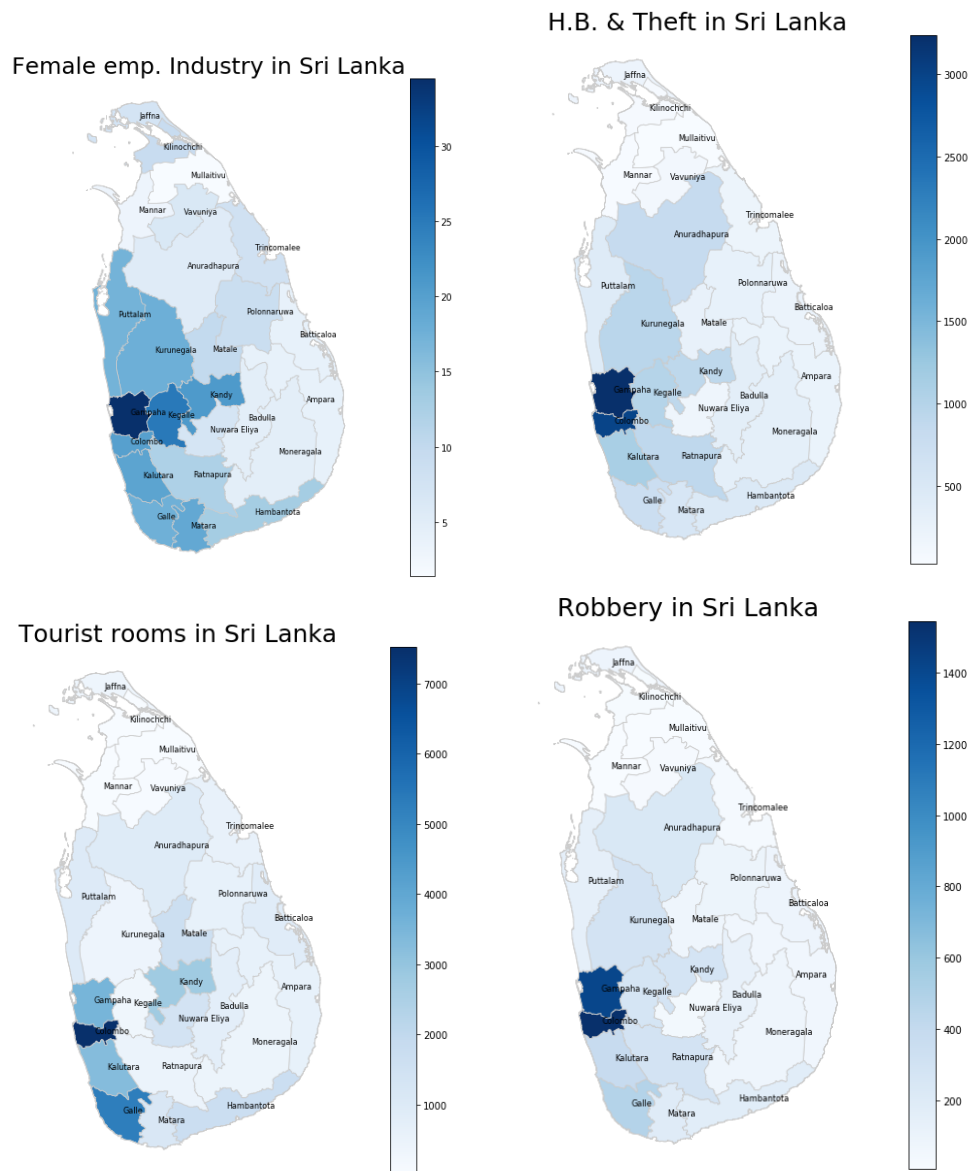


Fig. 1. Geo maps for crime

B. Methodology

We used the 'pandas' python framework to load and analyze data. Initially, data was loaded to separate data frames. Crime data and employment data were sorted on District and only the total room count per district was taken from the tourism data. The population of each district was also taken into a dataframe and crime counts were normalized by population. All these data frames were integrated to one data frame using district as the common factor since that was available on each data set to prepare a final integrated data frame.

Room count correlation(Pearson) was calculated and heat map was created for correlation coefficient which is greater than 0.6. "Fig. 2" shows that room count shows a higher correlation with almost all crime data. Robbery and Offence under Dangerous drugs are highly correlated. "Fig. 4" shows

scatter plots of each crime and tourism room count. Popular tourism districts such as Colombo, Galle, Gampaha, Kandy, Kalutara districts have recorded higher crimes as well. Each of these districts has more than 2000 tourist rooms, meaning that they are popular tourist destinations. Two hypotheses are defined with the tourism data. Hypothesis 1

- H0: Number of tourist Rooms and Crimes under Robberies are Uncorrelated
- H1: Number of tourist Rooms and Crimes under Robberies are Correlated

Hypothesis 2

- H0: Number of tourist Rooms and Crimes under "Offence under Poisons, Opium and Dangerous Drugs" are Uncorrelated

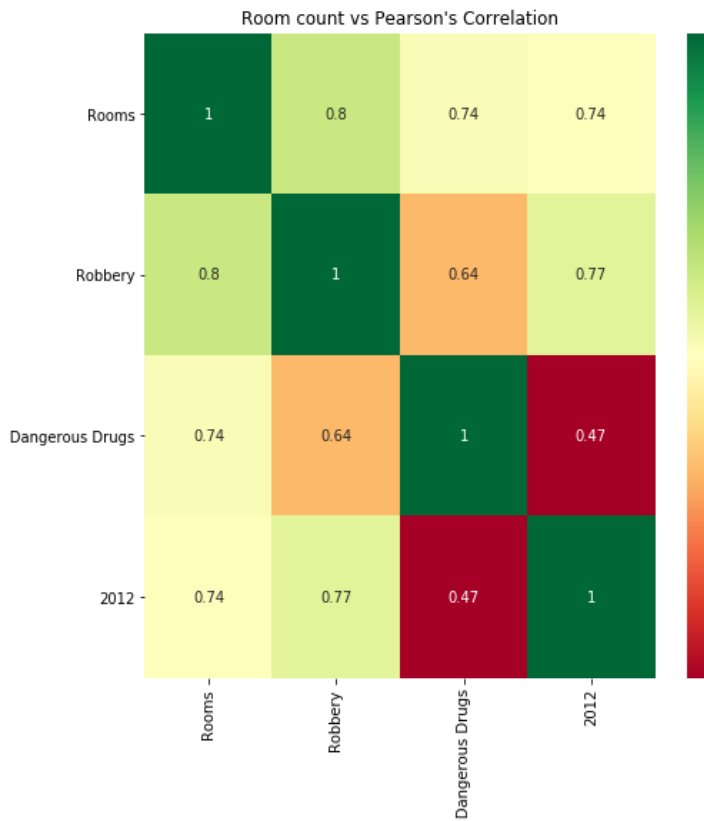


Fig. 2. Heat map of room count with crimes.

- H1: Number of tourist Rooms and Crimes under "Offence under Poisons, Opium and Dangerous Drugs" are Correlated

Before doing hypothesis testing we did a normality test on our explanatory variable room count and the response variables, robbery and offence under dangerous drugs. We could visualize that though there is a clear correlation between these variables. Although none of the variables were normally distributed. Therefore in order to quantify the correlation, Spearman's Rank correlation was calculated for above two hypotheses. Spearman rank-order correlation is chosen as the statistical procedure that is designed to measure the relationship between two variables where their distribution is unknown. With the results in both cases, we were able to reject the null hypotheses(H0) claiming that there is no correlation between above parameters with 95% confidence. A higher correlation was observed between the data for female employment in industrial sector and several categories of crimes such as abduction/kidnapping, Home break and theft, and Hurt by knife "Fig. 5". Normal distribution test was performed on the above crime data such as abduction/kidnapping, Home break and theft and Hurt by knife and all of them were found to be normally distributed. Therefore below hypotheses were tested for statistical significance with above data. Hypothesis 1

- H0: More Females working in Industrial sector has no relation to high Abduction / Kidnapping



Fig. 3. Heat map of female emp with crimes.

- H1: More Females working in Industrial sector has relation to high Abduction / Kidnapping

Hypothesis 2

- H0: More Females working in Industrial sector has no relation to high Hurt by Knife etc.
- H1: More Females working in Industrial sector has relation to high Hurt by Knife etc.

Hypothesis 3

- H0: More Females working in Industrial sector has no relation to high H.B. and Theft
- H1: More Females working in Industrial sector has relation to high H.B. and Theft

Before hypothesis analysis data set was split into two samples by female employment count. Low employment and high employment were used to test hypothesis testing. All null hypothesis were rejected by 0.05 significance level.

III. RESULTS

After analyzing relationship between crime and tourism/female employment were observed. Popularity of tourism causes a lot of crimes including,

- Robbery (0.81)
- Offence under Poisons, Opium Dangerous Drugs (0.74)

Female employment in industrial sector relates to crimes like,

- H.B. and Theft (0.75)

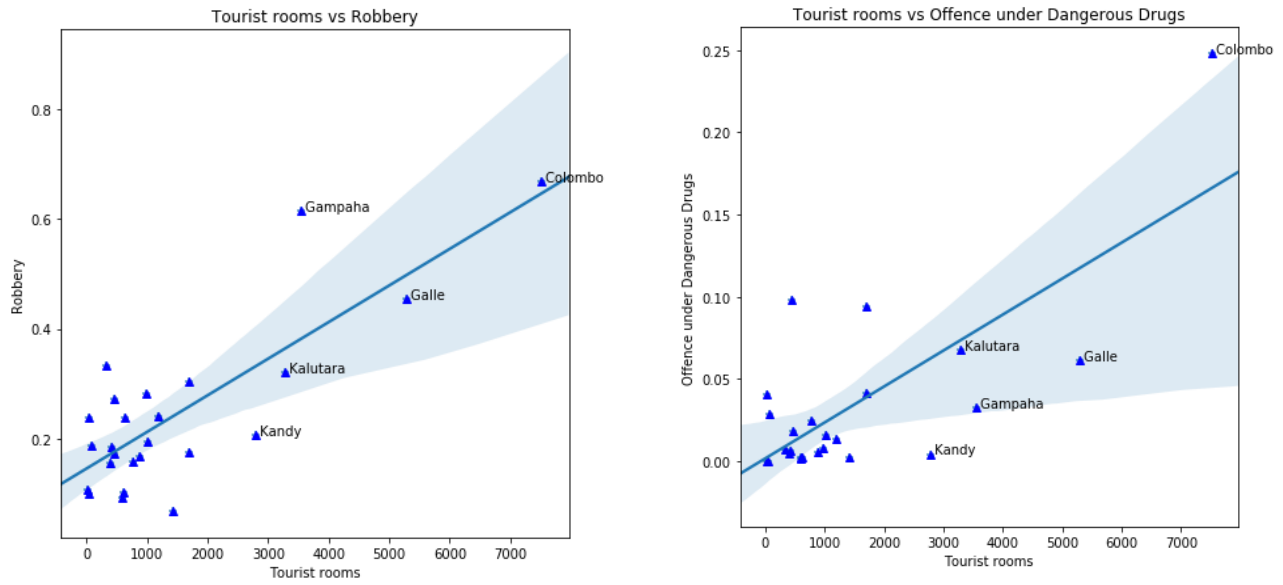


Fig. 4. Scatter plots for crime vs tourism rooms.

- Hurt by knife (0.83)

High female employment in industrial sector means high high criminal records. Refer scatter plots“Fig. 5”.

We use a linear model to predict H.B. theft given a district which has a particular female employment in industrial sector and/or tourism. Ridge regression was used since we have small data set (25 districts). Even though this is not a time-based prediction, still useful to predict theft for a district given female employment in industrial sector. Mean score of 3-fold cross validation

- Using female emp. (0.4913)
- Using tourism (0.0579)
- Using both features (0.5264)

Same technique was used for robbery predictions. Predicted robberies given a district with female employment in industrial sector and tourism popularity. Mean score of 3-fold cross validation

- Using tourism (0.4689)
- Using female emp. (0.1962)
- Using both features (0.6039)

IV. CONCLUSION

Tourism is often one of the most important driving forces behind the economy of given region, city or country. Tourist attraction is a blessing to the country as it contributes a major portion of the GDP of Sri Lanka. Though, according to our statistical results, there is a significant relationship with tourism and the number of crime categories such as robbery and Offence under Poisons, Opium Dangerous Drugs showing a higher rate in major tourist destinations. These criminal activities would affect the safety of tourists and the locals of that region. And the increased number of illegal drugs related crimes imposes a grave risk to the lives of the tourists and the

residents, which can possibly results in a massive breakthrough to the social and cultural values of the country affecting the future generations as well. In addition, increased number of crimes would result in a reduction of tourist attraction and therefore in a cascading manner it would affect the reputation of the country and the economy as well.

Having the above proven results, with a confident of 95% the high correlation between tourism and crimes such as robbery and Offence under Poisons, Opium Dangerous Drugs, it is highly advised that the necessary precautionary measures should be taken in order to avoid or minimize those crimes specifically. Sufficient security facilities should be provided to the tourists and additional security officials need to be employed at popular tourism destinations and the locals should be made aware to be more vigilant about the suspicious people or activities in the surrounding and both the tourists and the locals need to be encouraged to escalate such situations to the relevant officials as soon as possible.

We were also able to prove with a 95% confident that there is a high correlation between female employment in industrial sector and the crime categories such as Hurt by Knife and home break and theft. We ambition this nature is due to the complications that can possibly arise when more females are employed out from their homes at industrial sectors and are undergoing many ill social affairs causing the rise of above crimes such as hurt by knife. The increase in house breaks and theft is possibly because of the idling of houses when even the females are employed out at industrial sectors. In order to mitigate these issues, we encourage the relevant officials to spread this news to the public and to request to take more care to their loved ones working at industrial sectors and be vigilant to the possible risks to their lives and escalate them to the police as early as possible. People should be made that

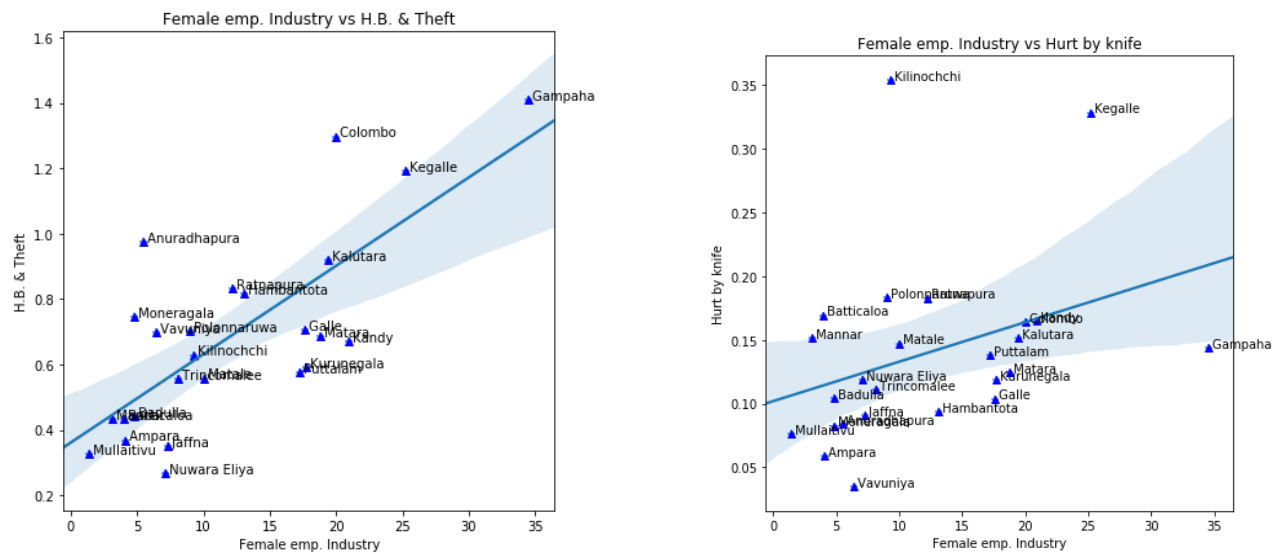


Fig. 5. scatter plot for crime vs female employment

they need to be more careful about the protection of their houses and properties when the females is also employed in industrial sector.

REFERENCES

- [1] Agnieszka Lisowska, "crime in tourism destinations: research review," Agnieszka Lisowska University of Wroclaw, Institute of Geography and Regional Development,.
- [2] open data portal, "<http://www.data.gov.lk/>".
- [3] Discover the beauty of LaTeX , "<https://www.latex-tutorial.com/>" .
- [4] Correlation Calculation "<https://machinelearningmastery.com/how-to-calculate-nonparametric-rank-correlation-in-python/>"