



جامعة مصر للمعلوماتية
EGYPT UNIVERSITY
OF INFORMATICS

Egypt University of Informatics
Computer and Information Systems
Data Analysis Course

The Analysis of Healthcare Dynamics in Egypt

Submitted by: Youssef Elmenshawy, Badr Mohamed, Omar Khaled, Ali Wael, Youssef Yasser

Introduction:

Healthcare in Egypt is characterized by a diverse landscape of medical services, each with its own set of fees, waiting times, and quality ratings. This report examines these variations to uncover patterns that could impact patient and provider decisions. By exploring how fees differ among specializations, the relationship between doctor ratings and appointment waiting times, and whether clinic fees influence doctors' ratings, this study aims to provide valuable insights into the Egyptian healthcare system. Understanding these factors can help improve patient experiences and guide policy changes to enhance the quality of care.

Research Question:

- What's the Key factors that affect the fees of doctors in Egypt?

Hypothesis tests:

- 1- Effect of specialization on fees
- 2- Fees and the doctor views
- 3- Fees and the waiting time
- 4- location and fees

Population of Interest:

Egyptian Doctors on a certain booking site

Sampling Method:

Simple Random Sampling

Bias Identification:

In designing this survey, we have taken steps to identify and minimize potential sources of bias. One potential bias we identified is that our sample is drawn exclusively from doctors who are listed on a booking website. To mitigate this bias, we ensured that our survey questions and analysis were focused on this specific group, acknowledging that our findings may not be representative of doctors who do not use booking websites.

Dataset:

We used a dataset from Kaggle of 1200 Egyptian doctors which are registered on a online booking site

We dropped a 330 from the dataset after filtering the null variables in specializations, average rating, waiting time, location, fees, rate count and doctor views.

Number of population: 1200

Number of samples used: 871.

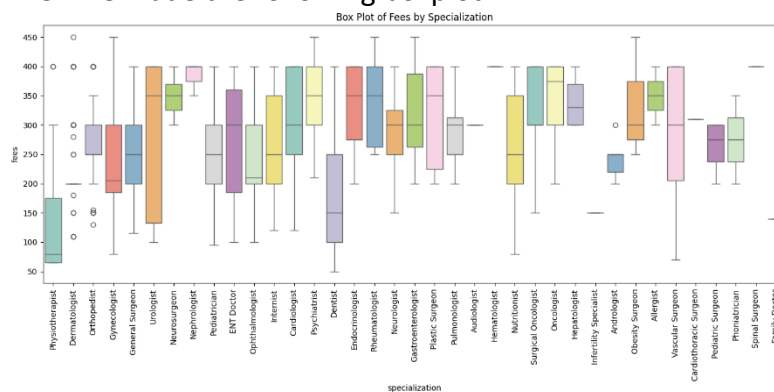
Analysis:

First Hypothesis:

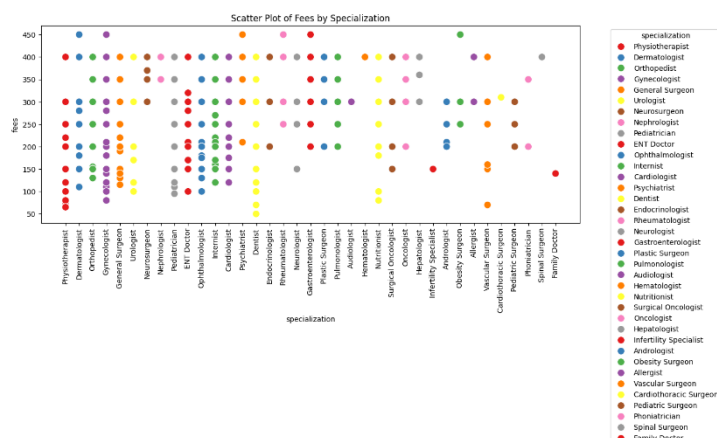
- Null Hypothesis (H0): there's no significant difference in fees across different specializations
- Alternative Hypothesis(H1): there's a significant difference in fees across different specializations

After doing the ANOVA test the p value turned to be less than 0.05 so we rejected the null hypothesis

Then we made the following boxplot:



And the scatter plot:



As you could see there's significant difference between the mean, median, and mode in different specializations compared to fees

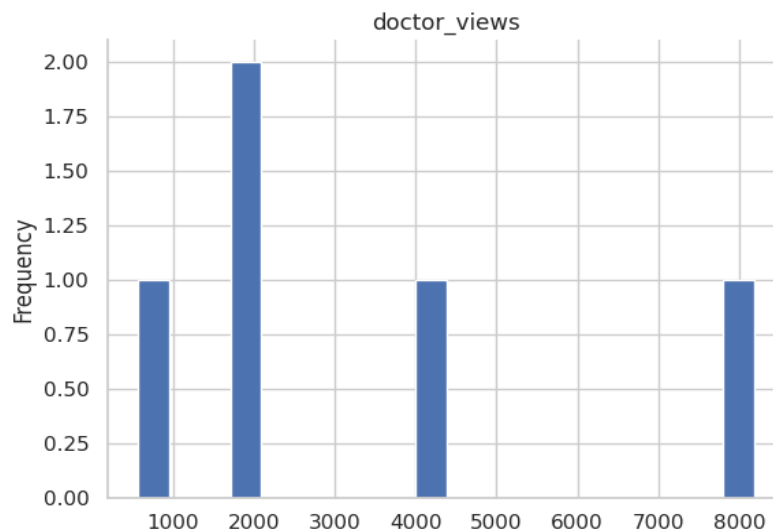
We also found that the top 3 specializations which charge the most are Neurosurgeon, Haematologist and Spinal Surgeon and the top 3 which charge the least are Physiotherapist, Family Doctor and Infertility Specialist

Second Hypothesis:

- first we created the graph for the fees distribution

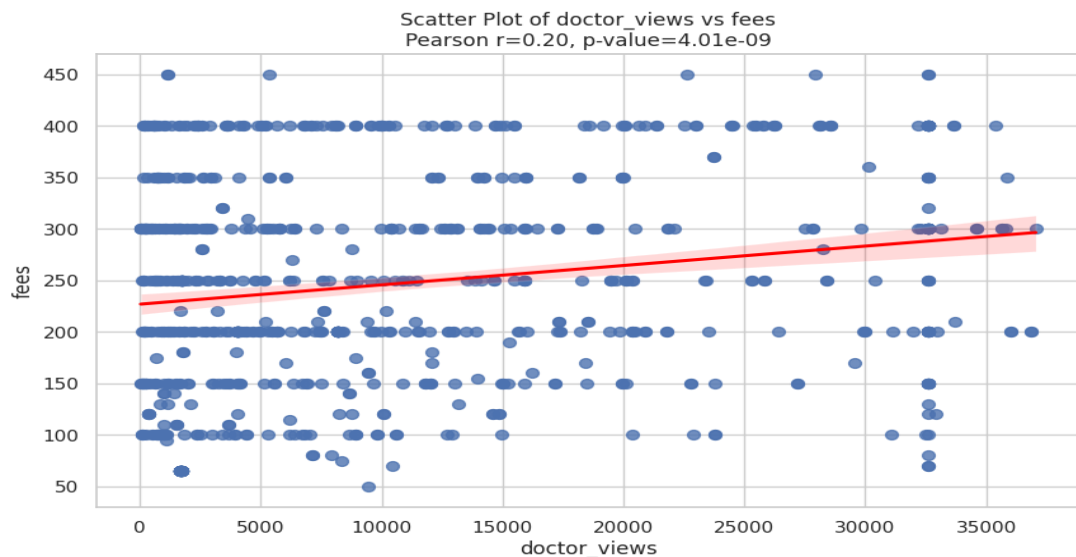


- then the graph for the doctor views distribution



- Null Hypothesis(H0): Pearson $r = 0$ that means that there is no relation between the fees and the doctor views
- Alternative Hypothesis(H1): Pearson $r \neq 0$ that means that there is a relation between the fees and the doctor views (non-directional relation)
- We used Pearson r to identify if there is a relation between the fees and the doctor views and after that we calculated the p value to see if the result is statistically significant and reject the null hypothesis.

- Pearson correlation coefficient: 0.19768085354297604 and since it is positive that means that there is a positive relation between the doctor views and the fees.
P-value: 4.007676045122018e-09 is smaller than 0.05 and that mean that we will reject the null hypothesis.



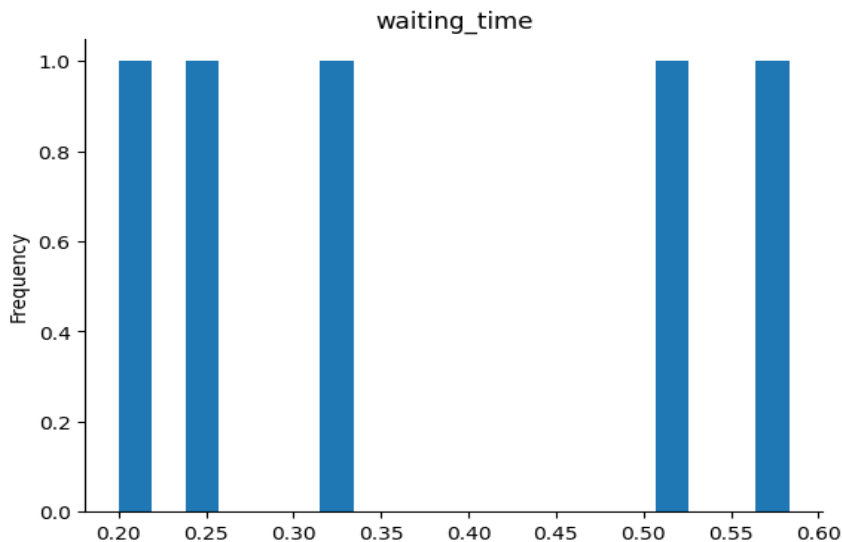
- This is a regression line that show the relation between the fees and the doctor views.

Third Hypothesis:

- first we created the graph for the fees distribution



- then the graph for the waiting time distribution



- Null Hypothesis(H_0): Pearson $r = 0$ that means that there is no relation between the fees and the waiting time
- Alternative Hypothesis(H_1): Pearson $r \neq 0$ that means that there is a relation between the fees and the waiting time (non-directional relation)
- We used Pearson r to identify if there is a relation between the fees and the doctor views and after that we calculated the p value to see if the result is statistically significant and reject the null hypothesis.
- Pearson correlation coefficient: 0.07502645417816026 and since it is positive that means that there is a positive relation between the doctor views and the fees.
P-value: 0.0268167145793388 is smaller than 0.05 and that means that we will reject the null hypothesis.

- This is a regression line that shows the relation between the fees and the doctor views.



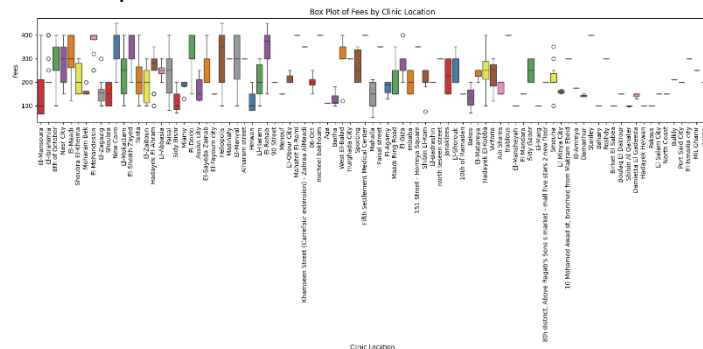
The location of the doctor's clinic does influence the fees.

-Null hypothesis(H_0): the location doesn't affect the fees

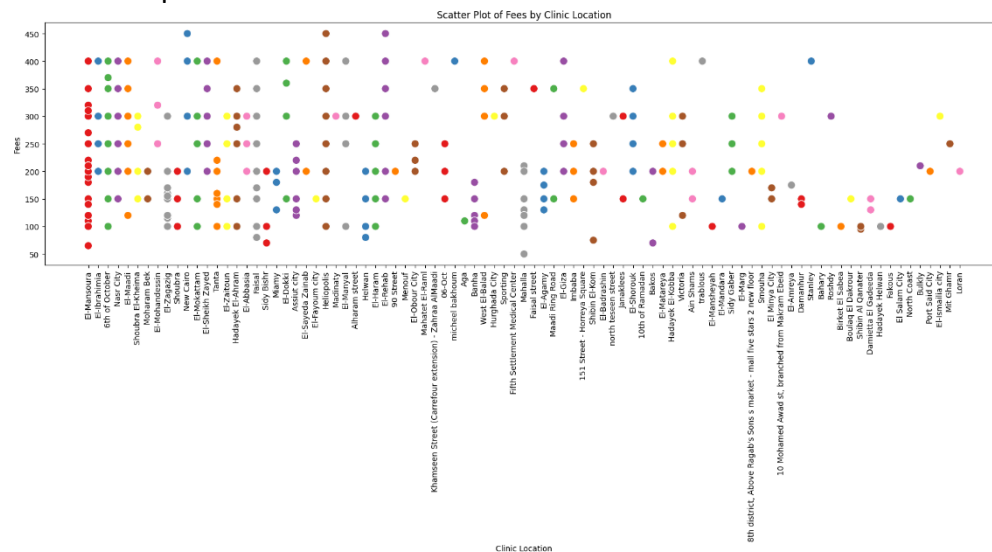
-alt hypothesis(H1): the location affect the fees

After calculating the P-value from the anove test it turned out to be less than 0.05 so we rejected the null hypothesis and after getting the box plot and scatter plot it turned out that there's significant difference between locarions and fees

The box plot:



The scatter plot:



Hypothesis Testing Steps:

- Step 1: define null and alternative hypotheses:
- First Hypothesis:
 - -Null Hypothesis (H_0): there's no significant difference in fees across different specializations
 - -Alternative Hypothesis (H_1): there's a significant difference in fees across different specializations
- Second hypothesis:
 - -Null Hypothesis (H_0): Pearson $r = 0$ that means that there is no relation between the fees and the doctor views
 - -Alternative Hypothesis (H_1): Pearson $r \neq 0$ that means that there is a relation between the fees and the doctor views (non-directional relation)

- Third Hypothesis:
- Null Hypothesis(H0): Pearson $r = 0$ that means that there is no relation between the fees and the waiting time
- Alternative Hypothesis(H1): Pearson $r \neq 0$ that means that there is a relation between the fees and the waiting time (non-directional relation)
- Fourth hypothesis:
- -Null hypothesis(H0): The location of doctor's clinic does not affect the fees
- -alt hypothesis(H1): The location of the doctor's clinic affect the fees
- Step 2: choosing a appropriate test for each hypothesis:
- First Hypothesis: ANOVA test , boxplot and scatterplot
- Second Hypothesis: pearson r, scatterplot
- Third Hypothesis: pearson r ,scatterplot
- Fourth Hypothesis: ANOVA test , boxplot and scatterplot
- Step 3: calculate p-value:
- First Hypothesis: 1.1102230246251565e-16
- Second Hypothesis: 4.007676045122018e-09
- Third Hypothesis: 0.0268167145793388
- Fourth Hypothesis: 7.383402432330294e-66
- Step 4: determining statistical significance:
- $\alpha = 0.05$

Conclusion:

Our findings indicate that several factors significantly impact the fees charged by doctors in Egypt. Specialization, the number of views a doctor receives, waiting time, and clinic location all play crucial roles. Notably, doctors with specialized skills, higher visibility, longer waiting times, and clinics in certain locations tend to charge higher fees. These insights are based on a specific sample of doctors listed on a booking website, and while these results are significant, they may not represent the broader population of Egyptian doctors

Any potential issues

1. Sample Bias: Our sample is limited to doctors listed on a specific booking website, which may not be representative of all Egyptian doctors.
2. Data Quality: Filtering out 330 samples due to missing values may affect the robustness of our analysis.
3. Geographic Distribution: Unequal representation of geographic regions could impact findings related to location and fees.
4. Variable Interactions: Potential interactions between variables like specialization and location were not deeply explored, possibly oversimplifying results.
5. External Factors: healthcare policies, and insurance coverage were not considered, leading to an incomplete analysis.