|  |  |
| --- | --- |
| Airlines delay and passengers satisfaction | |
|  | |
| January 2023Data visualization | Manuel AcquistapaceMiro RavaStefano Billeter |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | Syllabus  1. Abstract 2. Foreword, preface, acknowledgement 3. Introduction 4. Data sources 5. Data pre-processing 6. Interface design 7. Data visualizations 8. Conclusions 9. Next steps 10. Sources | |  |
|  |  | |  |
|  | Abstract In the highly globalized environment where we live, delays and setbacks are a real problem. These can affect travelers of all around the world, whether they are tourists or businessmen as well as the airports-airlines framework.  This data visualization project, runned in an academical context, intend to verify if there is a correlation between flight delays and passengers’ satisfaction. FOREWORD, PREFACE, ACKNOWLEDGEMENTS This report is part of the deliverables of the lab project in “data visualization” (prof. Giovanni Profeta). In this paper we go through all the aspects our work, from where we started to where it could be further developed.  Students Manuel Acquistapace, Miro Rava and Stefano Billeter are the creators of the webpage, visualizations and report. All the three are currently enrolled in the second year SUPSI’s bachelor degree in “Data Science & AI”.  We would like to thank prof. Giovanni Profeta for having guided us through this semester in a very important subject. We will never forget your kindness and willingness and we will treasure your teachings forever. Introduction The research we conducted aims to check the existence of a correlation between objective data (delays) and subjective data (satisfaction), our  initial hypothesis being that it exists  and it is strong and positive.  We used three datasets we found  on Kaggle; these will be  explained and analyzed in the next  chapter.  In order to verify our hypothesis,  we proceeded with the creation of  five visualizations, according to what  we have seen during the Data Visualization module.  The visualizations are shown in a  specific order to provide the  audience particular information along the way. Data visualization protocols and descriptions are also available for the user. Data sources As aforementioned we gathered our datasets from kaggle.com and skytraxratings.com. These have been carefully selected in order to be able to retrieve the best possible information, from a quality and reliability point of view. Here a glimpse of our data:   1. Air travels (year 2009)   Because of its large number of  variables and expanse,  this dataset allowed us to plot  the first two visualizations.   1. Airports dataset   We needed this dataset to complete the first one. In fact, this was useful to get all the information about the airports and their coordinates.   1. Skyratings   With the third and last dataset, we are able to come to our conclusion, that we will see in the last chapter of this report.  It basically shows the passengers’ satisfaction (rating) on an airport basis. Data pre-processing In this part of our project, we applied some of the techniques learned during the past semesters in the programming and supervised learning module. We used python through the computing platform Jupyter, where we imported our dataset and implemented some feature engineering. Essentially, we merged, deleted, dropped and performed other operations on our data, conductive to put together the information that we needed without the burden of tons of other useless, at least for our purposes, data.  #CHIARIRE I PASSAGGI SVOLTI PER QUANTO ATTIENE I TRE DATASETS Interface design #CHIARIRE COME SI SONO CREATE LE VISUALIZATIONS, SOFTWARES, SITI, ECC…. DATA VISUALIZATIONS As prior mentioned, we were able to generate three visualizations. We will now go through each of them and provide the relative explanation.  1)  The visualization shown in the map called “FLIGHT ROUTES” represents the airline traffic with respect to each air route in the USA.  Its features and visual variables are:  Yellow dots (shape and colour): depict the airports locations. The bigger the point the higher its relevance (total number of flights).  Thickness of segments (size, orientation): the thicker the segment, the bigger the number of flights in that direction.  Color of segment (color): doesn’t have a precise meaning. It is just a palette of colors that helps the user to  understand the departure and arrival airport.  The goal of this visualization is to learn the distribution of the flights contemplated in our dataset.  There is the chance to select, by clicking on the map or writing the airport code/name in the designated field, a single airport in order to show the specific visualization.  2)  The visualization shown in the second map “ROUTES DELAY” represents the air routes in the USA, divided by airport as the first one, and the delays. Its features and visual variables are:  Colour: the palette goes from green to red, where the first mean a low accrued delays on that route and the latter the opposite.  Orientation: clearly visualize the route direction.  This visualization acts as a mean to understand which routes are most often late, respectively on time.  3)  The third chart, named “DELAY INSIGHT” it is just an alternative visualization of what we’ve already shown in the previous one.  It is an interactive and 3D map, where the user can navigate around the globe (as we know at the moment only the USA territory is shown).  On a average delay time, this chart allows to pick a specific route, both from the left hand side menu and the  globe itself. We think that this provides  a much better and les confused understanding of the previous one.  Its features and visual variables are:  4) With the help of the fourth plot, we approach the final conclusions of our project. “USER RATINGS” tells us the mean satisfaction ranking per each of the US airport we took into consideration.  Rankings are scaled 1 to 10, from red to green.  The usage of color and size is important to give to the user, at a glance, a full and easy understanding of the map.  As for the earlier ones, we arranged it to be interactive (by clicking on one dot some information are automatically shown).  5)  The last, but not least, diagram is a correlation plot, which compares rankings and delays.  We think that plot nr. 4 is perfect to have a 365° understanding of the ranking along the USA, but we wanted to help the user with a more tabular and classical comparison.  The data shown in this visualization also confirms all the previous ones. CONCLUSIONS The findings in our research clearly state that there is a positive and strong correlation between flights delays and passengers’ satisfaction.  As stated in the abstract of the present report, in the kind of world ad environment we live today, we are pretty much influenced by delays and setbacks. It is interesting to notice how much these perturbations have an impact on the enjoyableness of our trip, probably more than the food we get on the airplane or the kindness of the flight assistants. We think we have been able, through the visualizations we created, to get the message across. NEXT STEPS These are few inputs that could be used to improve what we have done:  1) Expand the analysis and visualizations outside the US;  2) Compare other means of transport, like trains or buses.  3)Add other variables to the analysis, like the reason of the delay to see is there is any change in the perception of the setback. SOURCES  * Course slides and lectures * Report template: https://templates.office.com/en-us/papers-and-reports   edddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddeeeeeeeeeeeeeeeeeeeeeeDCdcWEceececeeeeeeCVFEVBRGBV |  |  |
|  | Image | |  |

|  |  |
| --- | --- |
|  |  |
|  |  |