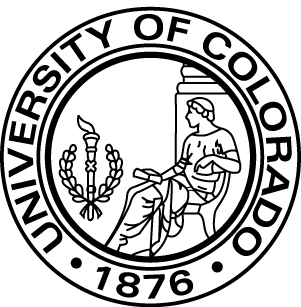
University of Colorado Denver



**Data Analytics for Environmental Data**

**Project Report**

Supervised by **Professor Farnoush Banaei-Kashani**

|  |
| --- |
| Team #9: |
|  |
|  |

Sai Vasudha Koya

Rania Abdraba

Somaia Alhazmi

Ramesh Malothu

**TABLE OF CONTENTS**

[Abstract 3](#_Toc418234039)

[A. Introduction 3](#_Toc418234040)

[B. Related Works 4](#_Toc418234041)

[C. Problem Definition 5](#_Toc418234042)

[D. Solution 7](#_Toc418234043)

[E. Results 8](#_Toc418234046)

[F. Conclusions 13](#_Toc418234049)

[G. Future Works 13](#_Toc418234050)

[References 14](#_Toc418234051)

**Abstract**

Our project idea is about Data analytics for Environmental data. Weather is an integral part of our lives. Weather is not something to be taken lightly or cheaply. Time series analysis models with explanatory variables that measure the weather quantitatively were used and applied to aggregate datasets of injury accidents. The main results reveal significant correlations on a yearly basis between weather variables and the aggregate number of injury accidents. Furthermore, we used the Regression Analysis Model to predict how weather condition impacts the number of accidents in the future. The outlook for the approach and its added value are discussed in the conclusion.

**Introduction**

As humans, the environment-the Earth-is our home. It is where we live, breathe, eat, raise our children, etc. Weather plays an important role in our environment. In most cases, recognizing the importance of a reliable weather service can produce positive results and help avoid negative consequences(accidents).

Weather conditions are considered to be a factor that affects the number of road accidents. Moreover, as the weather also affects mobility, it is to be expected that the effects of weather on the number of injury accidents and casualties are partly due to the changes in mobility occurring at the same time.

The aim of this project was to collect dataset about *Environmental data(weather condition and no of accidents),*and which illustrate the use of weather variables for analyzing changes in the number of road injury accidents. In our project, we collected dataset that describe number of accident for certain weather conditions, where we like to predict for future based on linear regression analysis.

Working of linear regression is explained in problem solution, by creating linear equations. This can be shown in a graph (both analysis and the prediction). The results obtained with no.of accidents data and extreme weather data are detailed in Section RESULT.

**Related Works**

# In our project, we created database and populated data into database by using **My SQL** **5.5** that is considered (as of July 2013) the world's second most widely used [relational database management system](http://en.wikipedia.org/wiki/Relational_database_management_system). MySQL is a popular choice for creating databases.

# Image result for mysql

# 

In addition, we created user interface by using **NetBeans IDE 7.2.1** which is a [software development](http://en.wikipedia.org/wiki/Software_development) platform written in [Java](http://en.wikipedia.org/wiki/Java_%28programming_language%29). The NetBeans [Platform](http://en.wikipedia.org/wiki/Platform_%28computing%29) allows applications to be developed from a set of modular [software components](http://en.wikipedia.org/wiki/Software_component) called modules. It consists of Java Script pages which is uses to program the behavior of web pages. We used it to create the project’s interface like **dataset**, **index**,…etc. It helped us to create a new web interface to show our prediction.



For prediction and analysis, we used Linear Regression Analysis. It consists of generating linear equation to predict for the future. Linear regression analysis is the most widely used of all statistical techniques: it is the study of *linear*, *additive* relationships between variables. Let Y denote the “**dependent**” variable whose values you wish to predict, and let X1, …,Xk denote the “**independent**” variables from which you wish to predict it, with the value of variable Xi in period t (or in row t of the data set) denoted by Xit.

# For generating the graph we used **Minitab software**. Minitab 17 offers a complete set of statistics in a user-friendly design that makes it easy to learn how to analyze data. It will be illustrated in Result Section.



**Problem Definition**

# In our project, we described how the weather condition affects the number of accident by increasing no of fatal, injury per year. We observed that weather condition brings more consequences than we expect for environmental problem. However, the more problems in world that encourage us to make solutions for these problems, to save our planet form any environmental changes.

# The idea of this project was about how we can choose environmental information, collect data about specific topic from the environment and to analyze this data and predict the changes that may cause in the future according to what we have already collection of information that has relation between the no of accident and weather condition

Depending on what we have mentioned previously, we collected information about the number of accidents that have happened over the past years starting (2001 to 2015), study the relationship between these numbers of accidents with the weather conditions. As a result of climate of our world varies by location and by time of year. Therefore many people interests to know what are the temperature and the state of weather that are constantly changing before leaving home to this matter of importance for them to take caution of the consequences of that. There are many weather conditions that affect and cause many serious accidents. For example, weather acts through visibility impairments, precipitation, high winds, and temperature extremes to affect driver capabilities, vehicle performance (i.e., traction, stability and maneuverability), pavement friction, roadway infrastructure, crash risk, traffic flow, and agency productivity.

Moreover, from the information that we collected in our database there are proportional relations between the number of accidents and the weather conditions. So, when we started in 2001 to 2003 the number of accidents increased when the weather increased as well. However, in 2005 demonstrates that the higher temperature was with less accident in these years. From 2006 until 2013 there are fluctuations between accident and weather conditions so on.

In order to know that an estimated prediction the number of accident by the increase or decrease in the coming years, The basic idea for prediction is to calculate the mean, but inorder to find an accurate and good approach to calculate this prediction we used the **Linear Regression Analysis**. Fortunately, we found the **Linear Regression Analysis** as new approach to calculate the prediction.

This is because models which depend linearly on unknown parameters are easier to fit than models which are non-linearly related to their parameters, because the statistical properties of the resulting estimators are almost accurate and easier to determine.

We have already collected this data, and have been configured and populated this data into database by using My SQL and connected it to the web pages by using Java Script to display the data and the results. Furthermore we used regression analysis model to offer expectation for the number of accidents, according to the weather changes in the future using the Minitab to draw linear graphs.

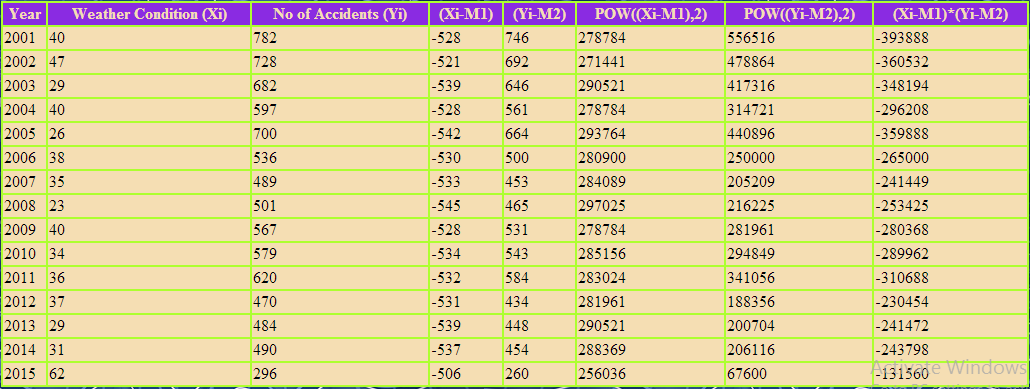
**Solution**

**Problem Statement**

We selected weather conditions and number of accidents from last 15 years, and how linear regression equation best predicts statistics performance.

**How to Find the Regression Equation**

In the table below, the xi column shows weather condition. Similarly, the yi column shows number of accidents. The last two rows show sums and mean accidents that we will use to conduct the regression analysis.



xi is the independent variable.

yi is the dependent variable.

The regression equation is a linear equation of the form: ŷ = b0 + b1x .

To conduct a regression analysis, we need to solve for b0 and b1. Computations are shown below.

b1 = Σ [ (xi – x^)(yi – y^) ] / Σ [ (xi – x^)2]  b0=  y^ - b1 \* x^

b1=-1.0018227 b0 =605.0353

Where x^ is mean of weather condition

Y^ is mean of number of accidents

Therefore, the regression equation is: Y=605.0353+-1.0018227X(X is the assumed weather condition)

**How to Use the Regression Equation**

Once you have the regression equation, using it is a snap. Choose a value for the independent variable (x), perform the computation, and you have an estimated value (ŷ) for the dependent variable.

In our project the independent variable is weather condition. The dependent variable is number of accidents.

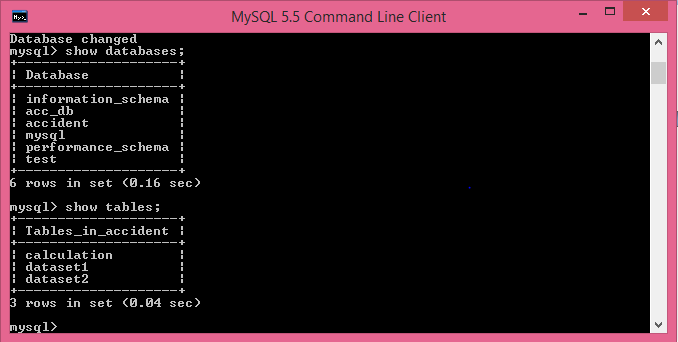
For the future if the weather condition is about 36 degree Fahrenheit then number of accidents in that year would be

605.0353+-1.0018227\*36= 568.9696

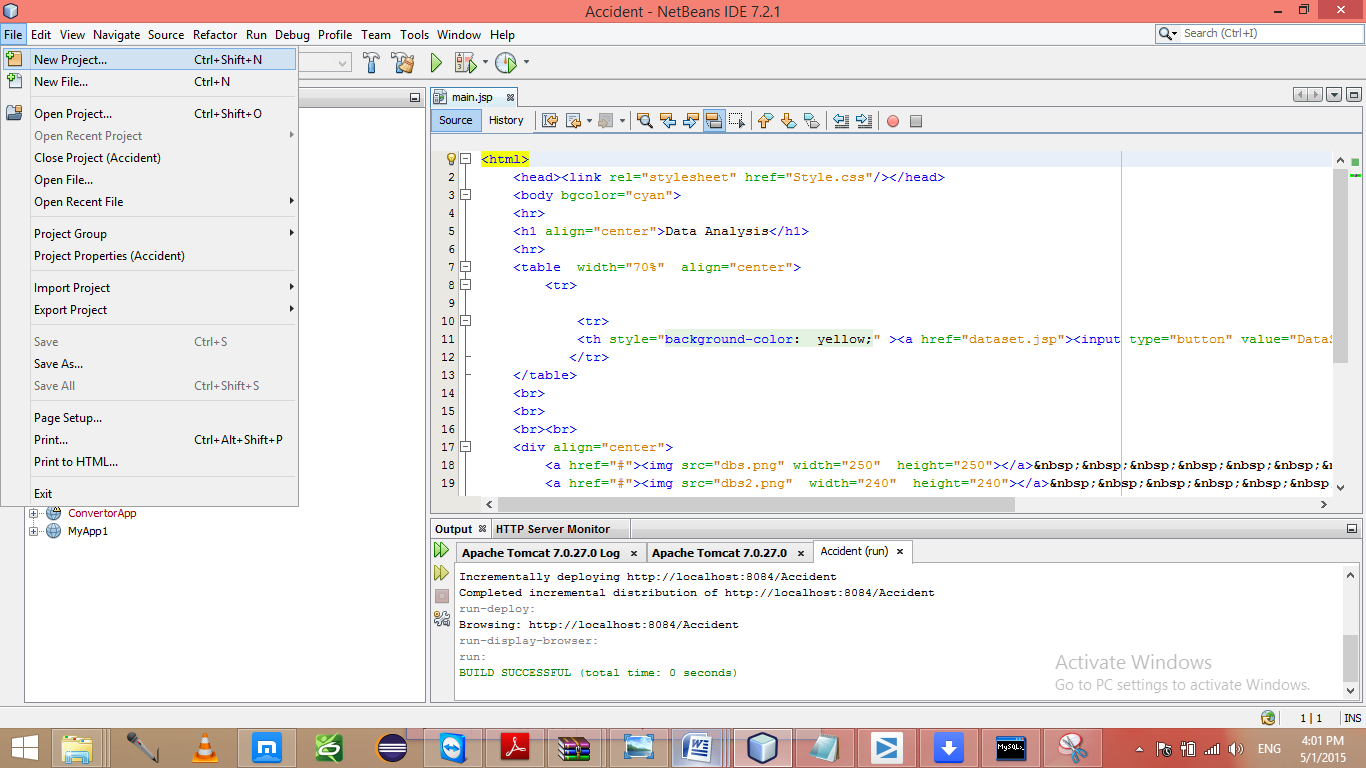
That’s how we predicted number of accidents for the future.

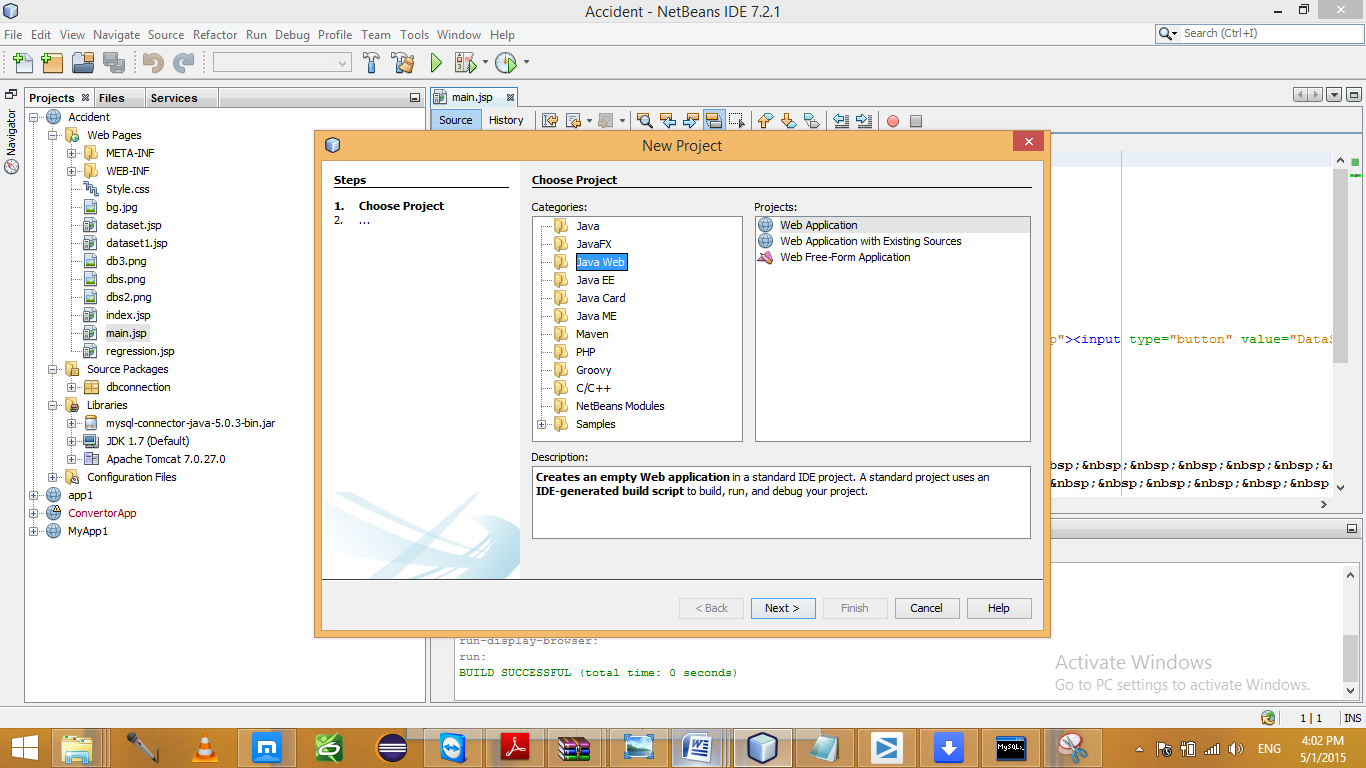
**Results:**

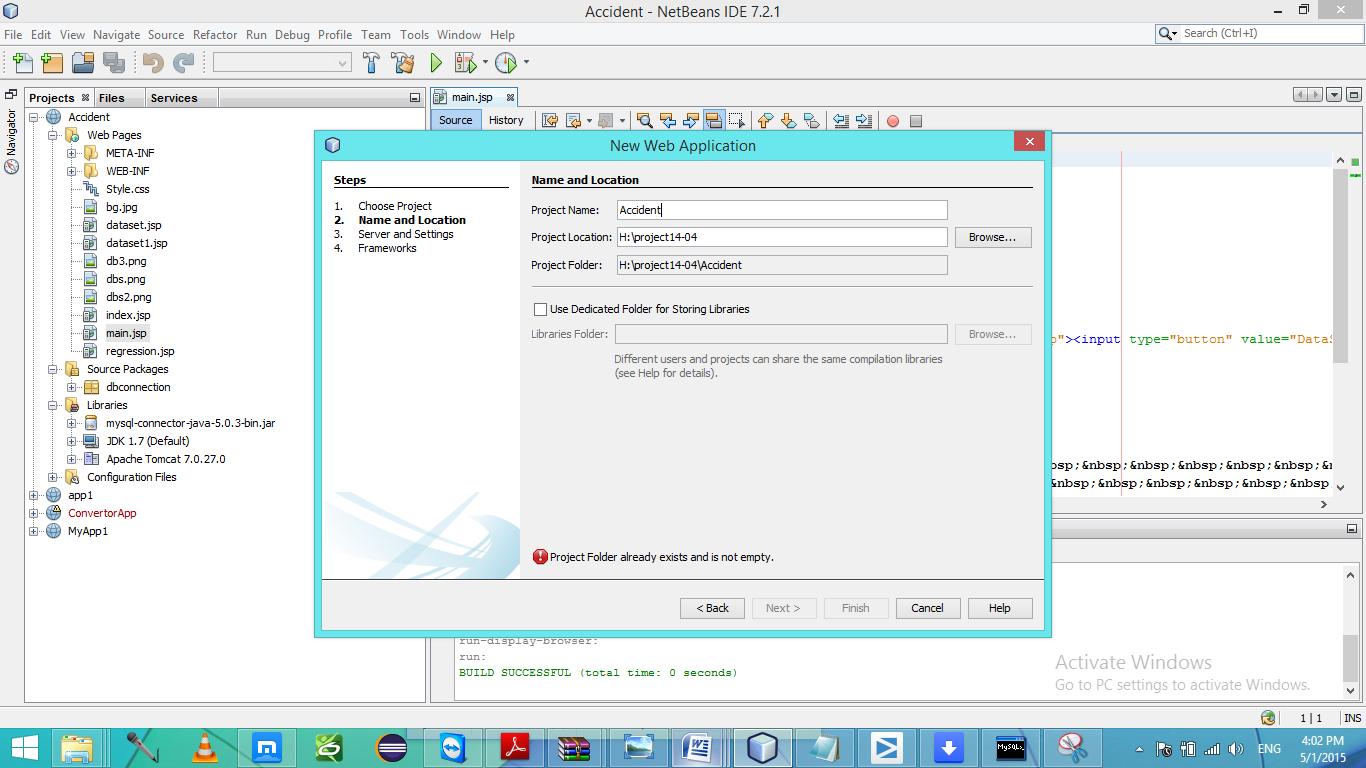
Our database accident contains three tables:



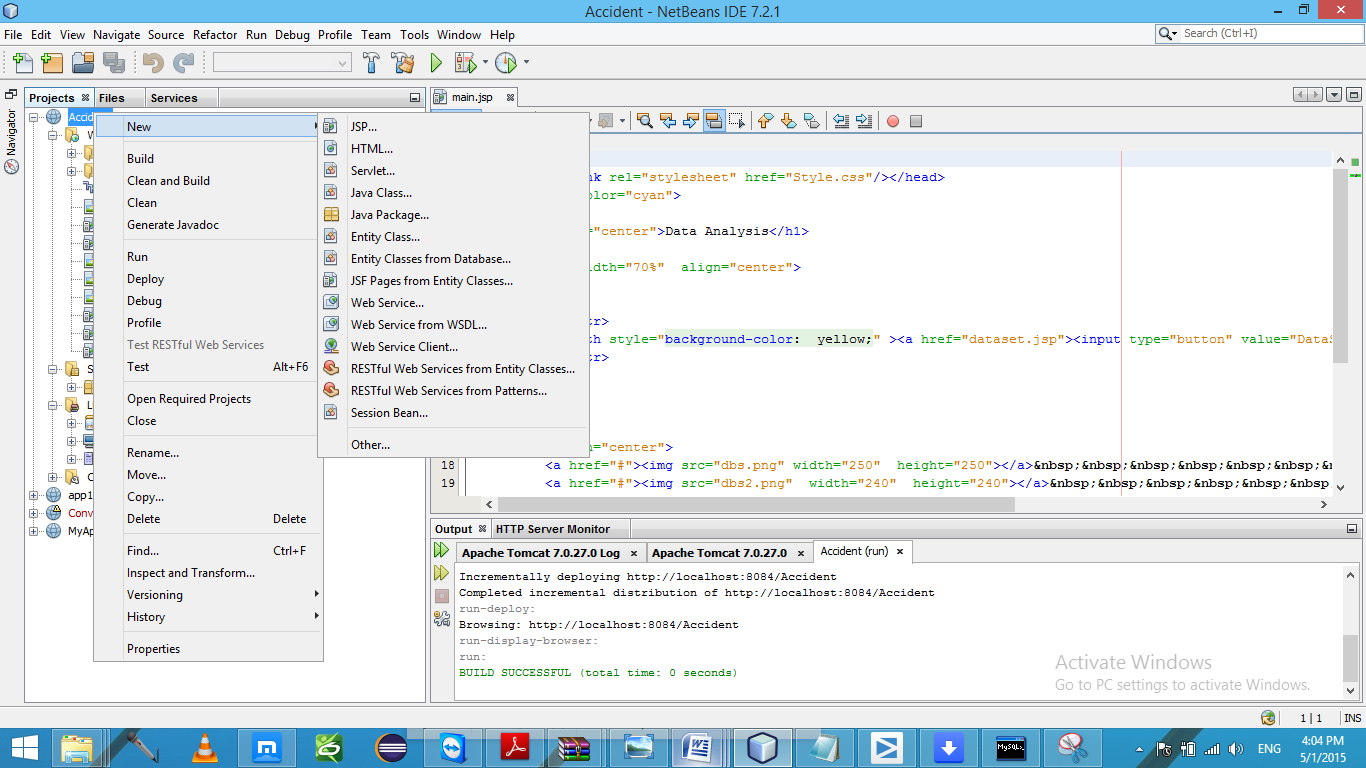
Create the project in net beans:

****

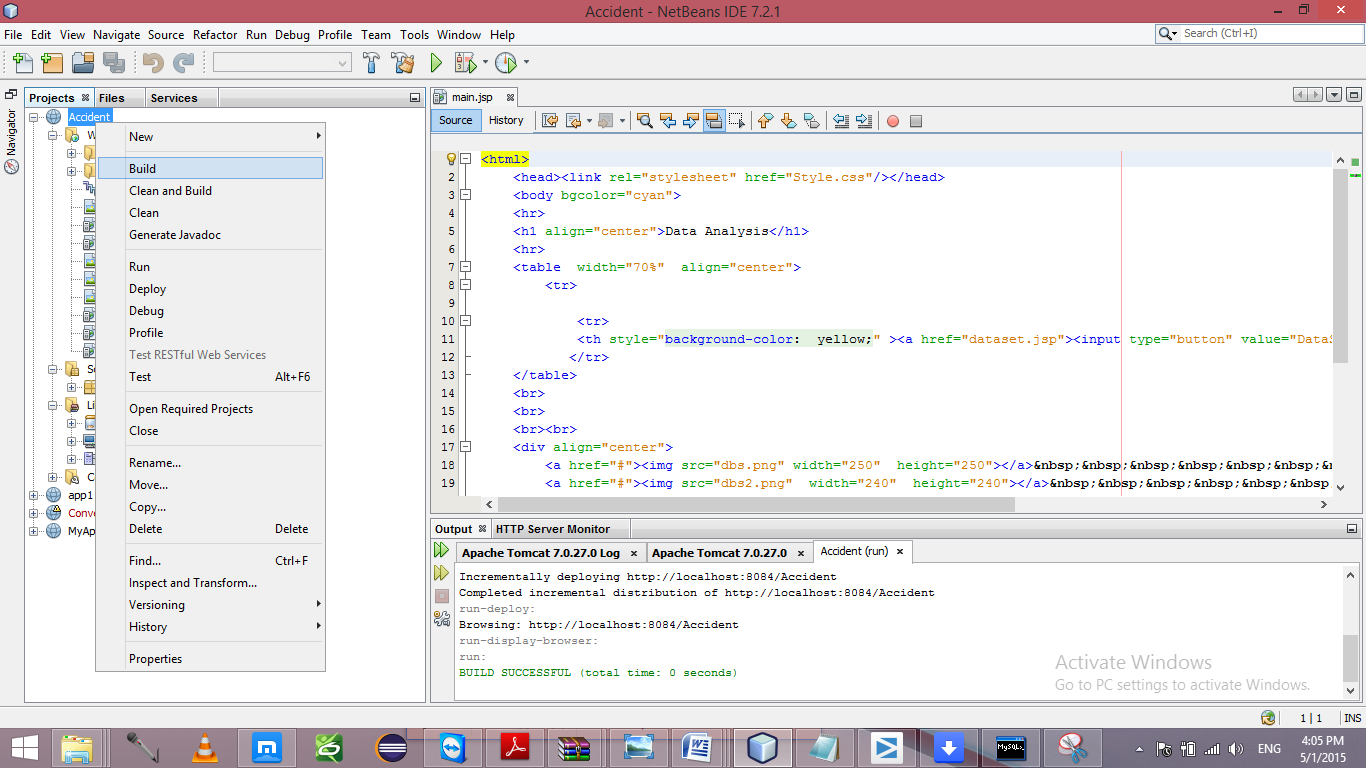
****

****

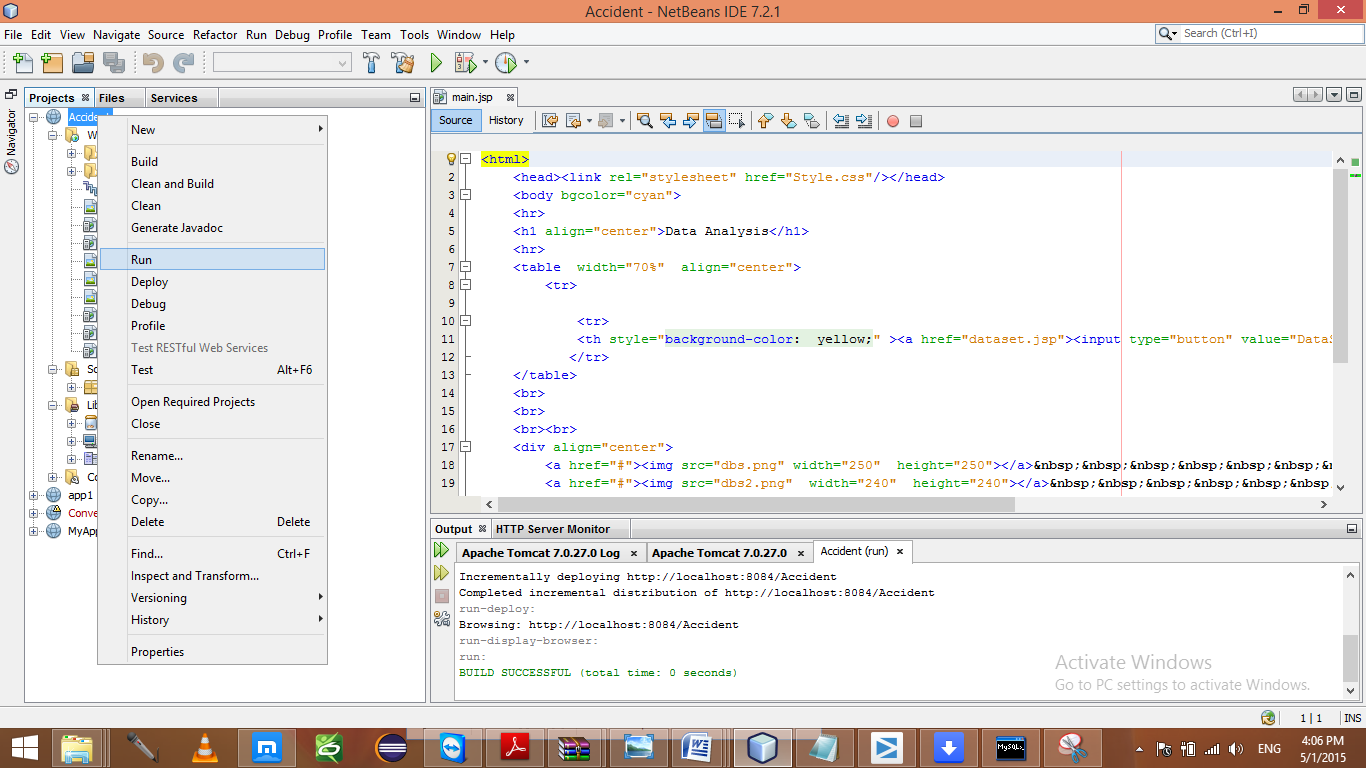
Create new jsp pages for user interfaces.

****

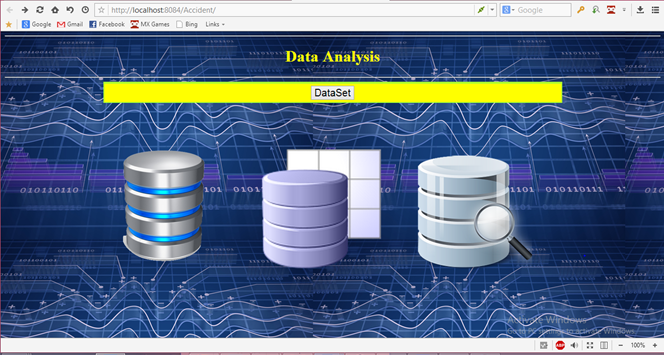
Build the project:

****

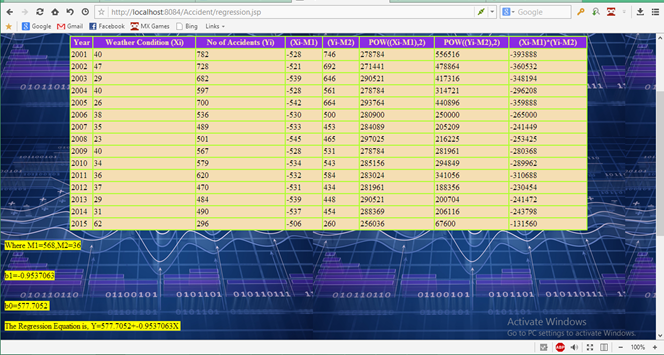
Run the project:

****

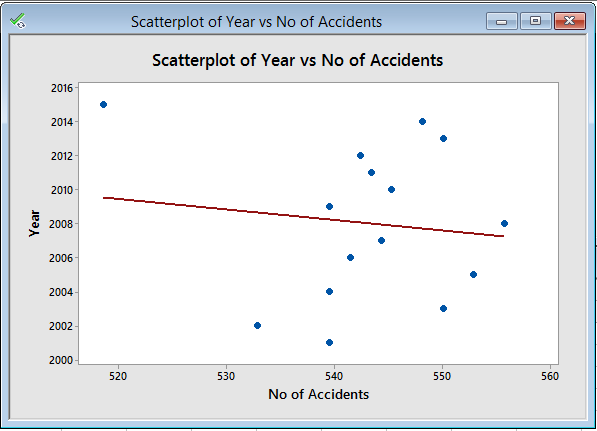
# **Web interface:**



After applying linear regression analysis:



An our final graph using minitab tool is like:



# Conclusion

This analysis helps for the prediction of number of accidents in the future. So that we can save number of lives on earth.

Future Work

We can increase number of variables i.e not only on single attribute weather but also on many other depending attributes such as drink and drive accidents, speed limits etc.

**References**

<http://blog.minitab.com/blog/adventures-in-statistics/regression-analysis-tutorial-and-examples>

<http://www.minitab.com/en-us/downloads/>

<http://people.duke.edu/~rnau/regintro.htm>

<http://en.wikipedia.org/wiki/Linear_regression>