Calidad del aire en Colombia 2011-2017

# Fase 1: Identificación del problema

## Descripción del contexto problemático (causas y síntomas):

Buscando tener un mayor control en el impacto medio ambiental que genera cada departamento del país, y crear un historial para analizar la evolución de la calidad del aire, esto último nos permite saber que zonas han sido las más afectadas y así tomar medidas preventivas para evitar llevar a una situación crítica a la biodiversidad local.

Uno de los mayores problemas que viven ahora los colombianos, así como el resto del mundo se debe a la elevada contaminación del aire provocada por múltiples sectores económicos como el agroindustrial o las fábricas de procesamiento de materiales con una gran huella de carbono. Precisamente aquí, en el Valle del Cauca, hay una gran zona industrial ubicada en Yumbo, como ciudadanos de Cali, queremos saber en qué estado se encuentra el aire que respiramos.

## Identificación de necesidades:

* Se necesita mostrar la información de la base de datos que el gobierno de Colombia nos facilita.
* Se necesita filtrar la información mostrada.
* Se necesita generar un reporte con los datos que se obtienen en todo el país.
* Se necesita generar un reporte con los datos que se obtienen en el Valle del Cauca.

## Definición del problema:

Actualmente, la información que suministra el gobierno de Colombia sobre la calidad del aire es obtenida por medio de terceros, pero no está claro cuál es el impacto en cada zona a lo largo del tiempo y tampoco se sabe si las mediciones se realizan en ubicaciones clave para que se represente con veracidad el estado actual del aire en nuestro país, para así decidir si se deben hacer cambios en el actual sistema de medición.

# Fase 2: Recopilación de la información necesaria

## Marco Teórico

### Información de las Columnas en el Conjunto de Datos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nombre de Columna** | **Nombre de Campo API** | **Descripción** | **Tipo** | **Tipo de Dato** |
| **Fecha** | fecha | Fecha en la cual se realizó la medición | Texto simple | String |
| **Autoridad Ambiental** | autoridad\_ambiental | Autoridad Ambiental en cuya jurisdicción se ubica la estación de monitoreo de la calidad del aire | Texto simple | String |
| **Nombre de la estación** | nombre\_de\_la\_estaci\_n | Nombre que identifica la estación de monitoreo de la calidad del aire | Texto simple | String |
| **Tecnología** | tecnolog\_a | Especifica la tecnología del equipo con el cual se realizó la medición | Texto simple | String |
| **Latitud** | latitud | Latitud de la ubicación de la estación | Número | Integer |
| **Longitud** | longitud | Longitud de la ubicación de la estación | Número | Integer |
| **Código del departamento** | c\_digo\_del\_departamento | Código DANE del departamento | Número | Integer |
| **Departamento** | departamento | Nombre del departamento | Texto simple | Integer |
| **Código del municipio** | c\_digo\_del\_municipio | Código DANE del municipio | Texto simple | Integer |
| **Nombre del municipio** | nombre\_del\_municipio | Nombre del municipio | Texto simple | String |
| **Tipo de estación** | tipo\_de\_estaci\_n | Clasificación de la estación según la duración del monitoreo | Texto simple | String |
| **Tiempo de exposición** | tiempo\_de\_exposici\_n | Periodo normativo en el cual se toma la medición | Número | Integer |
| **Variable** | variable | Parámetro climatológico o de contaminación ambiental evaluado por la estación de monitoreo de la calidad del aire | Texto simple | String |
| **Unidades** | unidades | Unidad de medida de la variable | Texto simple | String |
| **Concentración** | concentraci\_n | Valor obtenido durante la realización de la medición | Número | Integer |
| **Nueva columna georreferenciada** | geocoded\_column | Aparente posición desde la que se tomaron los datos | Ubicación | Location(Array)\* |

“Location” es un tipo que contiene latitud, longitud y una dirección. Se accede a “Location” como un arreglo con 2 elementos en él. Estos elementos son:

* La latitud de la localización. Puede ser nula, si el objeto ha sido recientemente añadido. Este debe ser de grado decimal, por ejemplo: 41.8657007325722
* La longitud de la localización. Puede ser nula, si el objeto ha sido recientemente añadido. Este debe ser de grado decimal, por ejemplo: -87.76110202195098

### SODA API

El Socrata Open Data API permite acceder mediante programación a una gran cantidad de recursos de datos abiertos de gobiernos, organizaciones sin fines de lucro y ONG de todo el mundo.

## Estado del Arte

Without data visualizations, drawing conclusions from large data sets or even discerning usable information from them is almost impossible. Using data visualization methods, designers can make information understandable for stakeholders.

Data visualization techniques include charts (line, bar, or pie), plots (bubble or scatter), diagrams, maps (heat maps, geographic maps, etc.), and matrices. There are many variations on these specific techniques that designers can use to meet specific visualization needs based on the data they’re working with.

Data visualization tools provide designers with an easier way to create visual representations of large data sets. When dealing with data sets that include hundreds of thousands or millions of data points, automating the process of creating a visualization makes a designer’s job significantly easier.

The best data visualization tools include Google Charts, Tableau, Grafana, Chartist.js, FusionCharts, Datawrapper, Infogram, ChartBlocks, and D3.js. The best tools offer a variety of visualization styles, are easy to use, and can handle large data sets.

### Tableau

Output options include multiple chart formats as well as mapping capability. That means designers can create color-coded maps that showcase geographically important data in a format that’s much easier to digest than a table or chart could ever be.

**Pros**

* Hundreds of data import options
* Mapping capability
* Free public version available
* Lots of video tutorials to walk you through how to use Tableau

**Cons**

* Non-free versions are expensive ($70/month/user for the Tableau Creator software)
* Public version doesn’t allow you to keep data analyses private

Tableau is a great option for those who need to create maps in addition to other types of charts. Tableau Public is also a great option for anyone who wants to create public-facing visualizations.

### Infogram

Infogram is a fully-featured drag-and-drop visualization tool that allows even non-designers to create effective visualizations of data for marketing reports, infographics, social media posts, maps, dashboards, and more.

Finished visualizations can be exported into a number of formats: .PNG, .JPG, .GIF, .PDF, and .HTML. Interactive visualizations are also possible, perfect for embedding into websites or apps. Infogram also offers a WordPress plugin that makes embedding visualizations even easier for WordPress users.

**Pros**

* Tiered pricing, including a free plan with basic features
* Includes 35+ chart types and 550+ map types
* Drag and drop editor
* API for importing additional data sources

**Cons**

* Significantly fewer built-in data sources than some other apps

Infogram is a great option for non-designers as well as designers. The drag-and-drop editor makes it easy to create professional-looking designs without a lot of visual design skill.

### ChartBlocks

ChartBlocks claims that data can be imported from “anywhere” using their API, including from live feeds. While they say that importing data from any source can be done in “just a few clicks,” it’s bound to be more complex than other apps that have automated modules or extensions for specific data sources.

The app allows for extensive customization of the final visualization created, and the chart building wizard helps users pick exactly the right data for their charts before importing the data.

Designers can create virtually any kind of chart, and the output is responsive—a big advantage for data visualization designers who want to embed charts into websites that are likely to be viewed on a variety of devices.

**Pros**

* Free and reasonably priced paid plans are available
* Easy to use wizard for importing the necessary data

**Cons**

* Unclear how robust their API is
* Doesn’t appear to have any mapping capability

#### AirQ+

Quantifying the effects of exposure to air pollution in terms of public health has become a critical component in policy discussion. WHO/Europe's software tool AirQ+ performs calculations that allow quantification of the health effects of exposure to air pollution, including estimates of the reduction in life expectancy.

AirQ+ estimates:

* the effects of short-term changes in air pollution (based on risk estimates from time-series studies);
* the effects of long-term exposures (using life-tables approach and based on risk estimates from cohort studies).

For each type of estimate, separate HELP files explain details of calculation.

Methodology and scientific basis for the risk estimates are summarized in the documents listed below.

# Fase 3: Búsqueda de soluciones creativas

## Ideas del procesamiento de la información:

## SODA API

Soda API allows us to interact with the data base by using a JSON file hosted in a unique URL that the API give us

* Text File

To this option, we have to download all the data base to access to it by reading line by line the text file

* SQL File

Saving the file as a sql file, and using the tools that C# give us to read the data base, to process all tha data needed

* SQL Server

Having the data base in a SQL server will allow us to do a search by the parameters that we want without needing to have the file in our computer

## Ideas para la creación de la visualización de los datos que el programa obtiene:

* Show one region air quality by a heat map

Having all the regions in a box in which the user can select which region want to see, and in a map will show how change the air quality in the time

* Show the country air quality by a heat map

Visualize all the country air quality in a map, and by a slider change the date of what you will see in the map

* Show one region air quality by pointing the stations

Pointing the station of one region can help us showing more detail in the information we want to show to the user and will save problems at the time of analize big quantity of data

* Show the country air quality by pointing the stations

Pointing the station of all the country will show more quantity of data for projects that might be need all this information

# Fase 4: Transición de la formulación de ideas a los diseños preliminares

## Ideas del procesamiento de la información:

## SODA API

Soda API can help us saving memory, and time at loading all the info in the software, but have a problem that is that the user will need internet to access to the information

* Text File

The text file will not be able to use just because of the data base size, the program will not be able to load all that data needed to process

* SQL File

SQL file will not be able to use in this project beacuse the file will have a really big size, which is not ideal if we want to save memory in the user computer

* SQL Server

This option won’t have the same problem as the SQL file, because we will have everything in the cloud and will just need internet conection to make everything work

## Ideas para la creación de la visualización de los datos que el programa obtiene:

* Show one region air quality by a heat map

This can be useful for working by separating the focus of the study and to do different things in each location, and is better because we don’t have to analize a big quantity of data

* Show the country air quality by a heat map

This can be useful for working on studies that focus on all the country, and analizing more data which is not recommended

* Show one region air quality by pointing the stations

This can be a good option beacuse the information will be more specific in the way of the data that will be shown is of an specific region, and in the way that the information of each station will be more specific too

* Show the country air quality by pointing the stations

This can’t be a good idea beacause the data that will be shown can be really heavy beacause of the quantity and how gruouped are

# Fase 5: Evaluación y selección de la mejor solución

* **Ideas del procesamiento de la información:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Easy to implement | Is ready to work | Efficiency showing the data | Total |
| SODA API | 4 | 5 | 5 | 14 |
| SQL Server | 3 | 1 | 5 | 9 |

1----------------------5

Worst-------------Better

The best idea is use the SODA API because the SQL server need to be assamble, while the SODA API is ready to use, being this, the crucial rule to the selection

* Ideas para la creación de la visualización de los datos que el programa obtiene:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Easy to implement | Efficiency to analize tha data | Nice at seen | Total |
| Show one region air quality by a heat map | 2 | 5 | 4 | 11 |
| Show one region air quality by pointing the stations | 3 | 3 | 2 | 8 |

1----------------------5

Worst-------------Better

The best option is to do a heat map using various parameters beacause will be easier to analize tha data, and at seen could less intrusive, and nicer