INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

ITAM

**MODELO PREDICTIVO DE NIVELES DE CONTAMINACIÓN**

TIPO DE TRABAJO:

APLICACIÓN PRÁCTICA

(ÁREA DE CONCENTRACIÓN: CIENCIAS DE DATOS)

TESINA

CON LA FINALIDAD DE OBTENER EL TÍTULO DE

INGENIERA EN NEGOCIOS

PRESENTA

VANESSA VELÁZQUEZ MOJICA

ASESOR: RAFAEL GREGORIO GAMBOA HIRALES

México, CDMX. 08/01/2018

CONTENIDO

1. ANTECEDENTES 4
2. PROBLEMÁTICA 6
3. PREGUNTAS DE TRABAJO 7
4. HIPÓTESIS DE TRABAJO 7
5. OBJETIVOS DEL PROYECTO 8
6. ESTRATEGIA METODOLÓGICA 9
7. REFERENCIAS BIBLIOGRÁFICAS PREVIAS 11
8. FUENTES SECUNDARIAS DE DATOS 12
9. TABLA DE CONTENIDO TENTATIVO DEL

REPORTE FINAL 12

1. CRONOGRAMA (GRÁFICA DE GANT) 12
2. CURRICULA DE LOS CONSULTORES 12
3. ANEXOS Y APÉNDICES: 12

MODELO PREDICTIVO DE NIVELES DE CONTAMIACIÓN EN LA CIUDAD DE MÉXICO

Vanessa Velázquez Mojica

**Resumen**

A través de éste proyecto de tesina se pretende analizar la información disponible sobre los niveles de contaminación ambiental en la Ciudad de México proporcionados por la Secretaria del Medio Ambiente del Distrito Federal (SEDEMA). La finalidad es desarrollar un modelo predictivo de los niveles de contaminación ambiental críticos con el objetivo de abrir paso a investigaciones más profundas en el tema y facilitar el desarrollo de mecanismos de control y prevención de dichos niveles críticos de contaminación ambiental. La relevancia de este trabajo radica en los siguientes puntos: primero; la Ciudad de México es una de las más pobladas del mundo haciendo susceptibles a severos problemas de salud pública asociados a la contaminación a una población de aproximadamente veintidós millones de habitantes. Segundo; los niveles de contaminación encontrados en la Ciudad de México son considerablemente mayores en comparación con los niveles de otras ciudades de tamaño y población similares[[1]](#footnote-1), lo que hace imperativo un cuestionamiento sobre las posibles causas a dichos resultados. Tercero, la Ciudad de México se encuentra dentro de los primeros diez lugares de las ciudades a nivel mundial[[2]](#footnote-2) que mayor invierten en protección ambiental y dicha inversión está demostrando ser poco eficiente con relación al segundo punto mencionado. En conclusión, dada la importancia de minimizar los riesgos en salud pública y efectos nocivos en el cambio climático surge este análisis para desarrollar un modelo predictivo de los niveles de contaminación críticos a través de métodos supervisados de aprendizaje de máquina.

**Palabras clave:** contaminación ambiental, aprendizaje de máquina, métodos supervisados, SEDEMA, modelos predictivos, SIMAT.

1. **ANTECEDENTES**

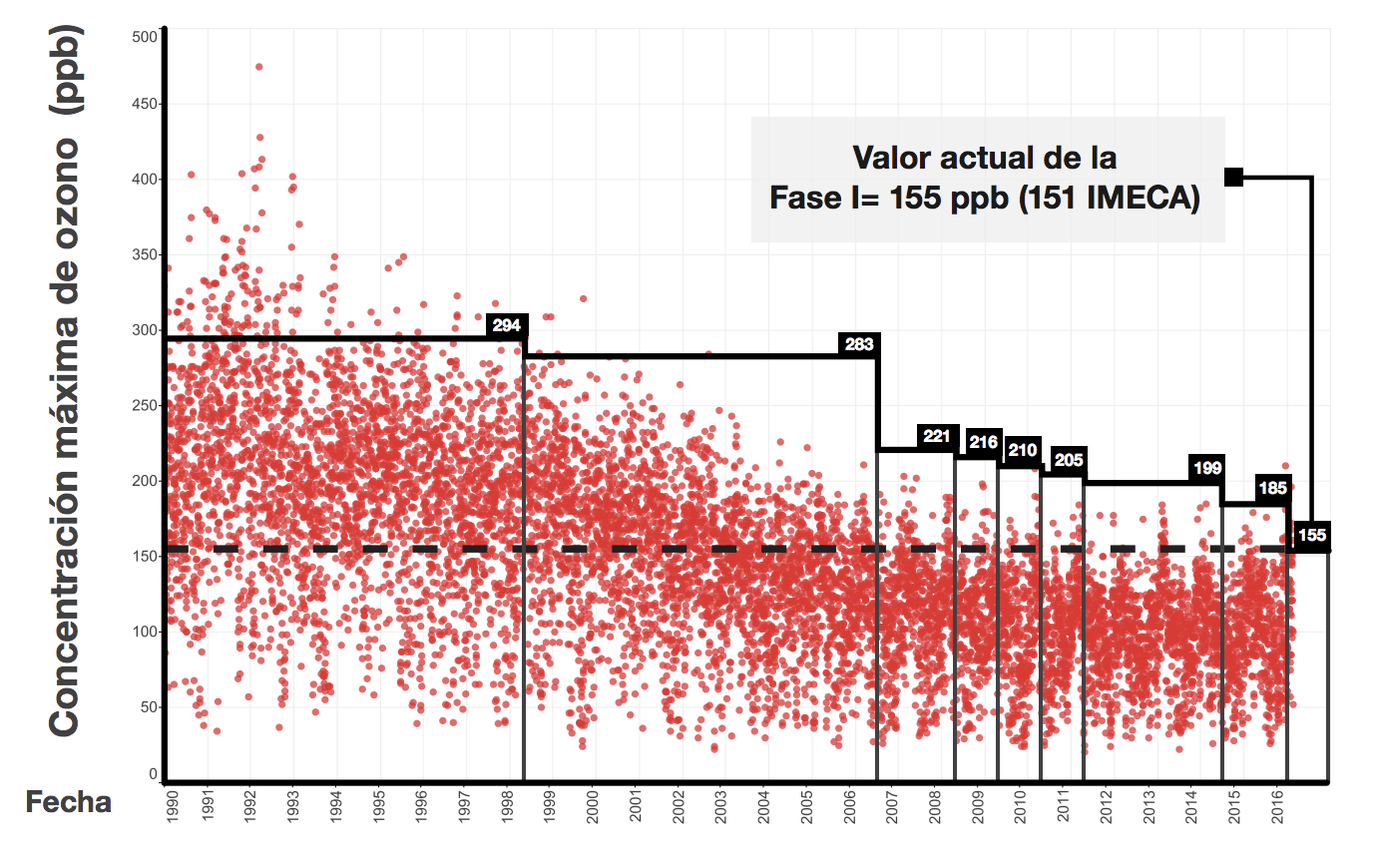
La Ciudad de México históricamente se ha consolidado como el punto de centralización del desarrollo económico y político del país, dicha centralización la coloca hoy en día como una de las ciudades más extensas en territorio con 9,560 km² y una de las más pobladas a nivel mundial con una población de aproximadamente 21,157,000 habitantes (en ambos casos considerando el área conurbada o metropolitana). Es así que la actividad humana de estos casi veintidós millones de habitantes que se desarrollan en la Megalópolis genera una gran cantidad y variedad de sustancias que modifican la composición natural del aire (Anexo 1). El deterioro de la calidad del aire por la presencia de estas sustancias contaminantes de origen antropomórfico (actividad humana) tiene un impacto nocivo en la salud pública y el medio ambiente. Esto es respaldado por diversos estudios en la Ciudad de México y otras ciudades con características similares que demuestran la existencia de una relación positiva entre el incremento de enfermedades cardiovasculares y respiratorias con el aumento de la concentración de contaminantes atmosféricos. Una manera en la que se ha buscado proteger la salud de la población es a través del monitoreo continuo de la calidad del aire que realiza el Sistema de Monitoreo Atmosférico (SIMAT) con los siguientes objetivos:

1. Vigilar el cumplimiento con las Normas Oficiales Mexicanas de salud ambiental
2. Evaluar la calidad del aire con respecto a indicadores de concentración de contaminantes que se establecen como criterio de comparación.
3. Cuantificar el nivel de contaminación ambiental al cual está expuesta la población
4. Mantener informada a la población sobre el nivel de contaminación ambiental actual y sus riesgos.
5. Generar información de manera inmediata y oportuna para la activación o desactivación de protocolos de emergencia derivados de una concentración de contaminación crítica nociva para la salud y el medio ambiente.
6. Informar de manera oportuna a la población sobre el estado de la calidad del aire.
7. Facilitar la evaluación de la distribución espacial y el transporte de los contaminantes atmosféricos a partir de la información recolectada de los centros de monitoreo.
8. Facilitar la evaluación y seguimiento de las estrategias para la gestión de la calidad del aire en la Ciudad de México y la zona conurbada.
9. Evaluar la tendencia histórica de los contaminantes

Para complementar el gran logro que representa contar con un sistema de monitoreo continuo de la calidad del aire bajo los objetivos previamente mencionados, hace falta implementar un modelo que realice un análisis robusto de la información recolectada por el SIMAT con el objetivo práctico de predecir los niveles de contaminación ambiental nocivos para la salud pública y medio ambiente, y que permita entender cuáles son los factores determinantes de los niveles críticos.

Es importante mencionar como antecedente que la SEDEMA utiliza actualmente el Índice Metropolitano de la Calidad del Aire (IMECA) como referencia en México para medir los valores de contaminación por zona dentro del área Metropolitana. Es a partir de este indicador que se toma la decisión de activar o desactivar los protocolos de emergencia ante una contingencia ambiental. Dicho valor de referencia inició en 1990 a partir de un valor de 294 partes por billón de ozono para activar la Fase I de contingencia, sin embargo, en el 2016 se modificó el estándar de la Fase 1 a 151 IMECA equivalente a 155 partes por billón de ozono

Gráfica 1. Fuente: SEDEMA, 2016.



(Los puntos indican la concentración máxima de ozono de cada día)

1. **PROBLEMÁTICA**

Para complementar el gran esfuerzo que SEDEMA y el SIMAT han aplicado en el monitoreo continuo y control de la calidad del aire de la Ciudad de México y la zona Metropolitana se propone en este proyecto de tesina el **desarrollo de un modelo predictivo de los niveles de contaminación a partir de la información recolectada por el SIMAT**. Dicho modelo, al ser de carácter predictivo puede complementar las actuales estrategias aplicadas para disminuir los efectos nocivos de los niveles críticos de contaminación ya que hasta el momento son protocolos de reacción, no así de prevención.

El problema consiste en encontrar el modelo que mejor ajuste las predicciones a las mediciones reales, encontrar cuáles son las variables que deben considerarse para el desarrollo y buen funcionamiento del modelo.

1. **PREGUNTAS DEL TRABAJO**

¿Cuál es la variable objetivo del modelo para predecir la calidad del aire?

¿De qué dependen los niveles de contaminación atmosférica o y la variable objetivo?

¿Existe alguna correlación entre las variables de monitoreo de calidad del aire del SIMAT y el índice de calidad del aire o la variable objetivo?

¿Existe alguna estacionalidad en los niveles de la variable objetivo?

¿Es cuantificable el impacto de los posibles factores de la variable objetivo en la contaminación ambiental?

¿Cuáles son los factores que se deben considerar para el desarrollo del modelo predictivo?

¿Cuál es el mejor método de predicción para desarrollar el modelo?

¿Cuál es el margen aceptable de error del modelo predictivo?

¿Cuál es la confiabilidad del modelo predictivo?

¿Cuáles son los costos asociados a las predicciones fallidas del modelo?

¿Cuáles son los beneficios económicos de la implementación del modelo?

1. **HIPOTESIS DEL TRABAJO**

A partir de las bases de datos del monitoreo continuo de la contaminación ambiental por parte de SEDEMA es posible desarrollar un modelo predictivo de los niveles de contaminantes del aire que permita hondar en la investigación sobre las causas de las crisis de la contaminación ambiental.

1. **OBJETIVOS DEL PROYECTO**

**El objetivo principal del proyecto es desarrollar un modelo predictivo de los niveles de contaminación en el aire a partir de los datos históricos de contaminación ambiental recolectados y proporcionados por el SIMAT.**

**5.1 PROBABLE IMPACTO DEL DESARROLLO DEL PROYECTO**

* Predecir de manera confiable los niveles críticos de contaminación ambiental
* A partir de la difusión de este modelo facilitar la investigación sobre las causas que influyen en las crisis de contaminación ambiental.
* A partir de la implementación del modelo de predicción facilitar la creación de políticas preventivas para evadir los puntos críticos nocivos de contaminación ambiental.
* Disminuir la variación en el índice de contaminación ambiental al implementar mayor control sobre los factores que la ocasionen a través de dichas políticas preventivas.
* Mejorar la calidad de vida de los habitantes de la Ciudad de México y la zona conurbada al minimizar los riesgos de salud y mortalidad ocasionados por la contaminación del aire.
* Disminuir los gastos de salud pública derivados de la atención que tiene que ser dirigida a los padecimientos ocasionados por la contaminación.

1. **ESTRATEGIA METODOLÓGICA**

Utilizando las mediciones en la base de datos de SEDEMA se desarrollará un análisis por estación de monitoreo (las estaciones se encuentran enlistadas en la tabla 1). Aunado a ese análisis por estación se hará un análisis a través de series de tiempo (los lapsos aún están por definirse). Para el desarrollo del modelo se tomará el catálogo de parámetros de medición de contaminación ambiental (tabla 2).

Se hará un primer acercamiento a los datos a través de un Data Profiling y un análisis exploratorio de los datos (EDA) para entender el tipo de datos que se van a manejar. A partir de estos procedimientos se obtendrán análisis gráficos entre las variables medidas y la variable objetivo que permitan estudiar la correlación entre las variables, su variación y hacer un análisis estadístico básico del cuál partirán las primeras preguntas y supuestos del modelo.

Posterior a dichas interpretaciones se presentarán alternativas para los modelos predictivos que mejor se adapten al tipo de datos y análisis que se busca realizar.

Tabla 1. Fuente: SEDEMA, 2017.

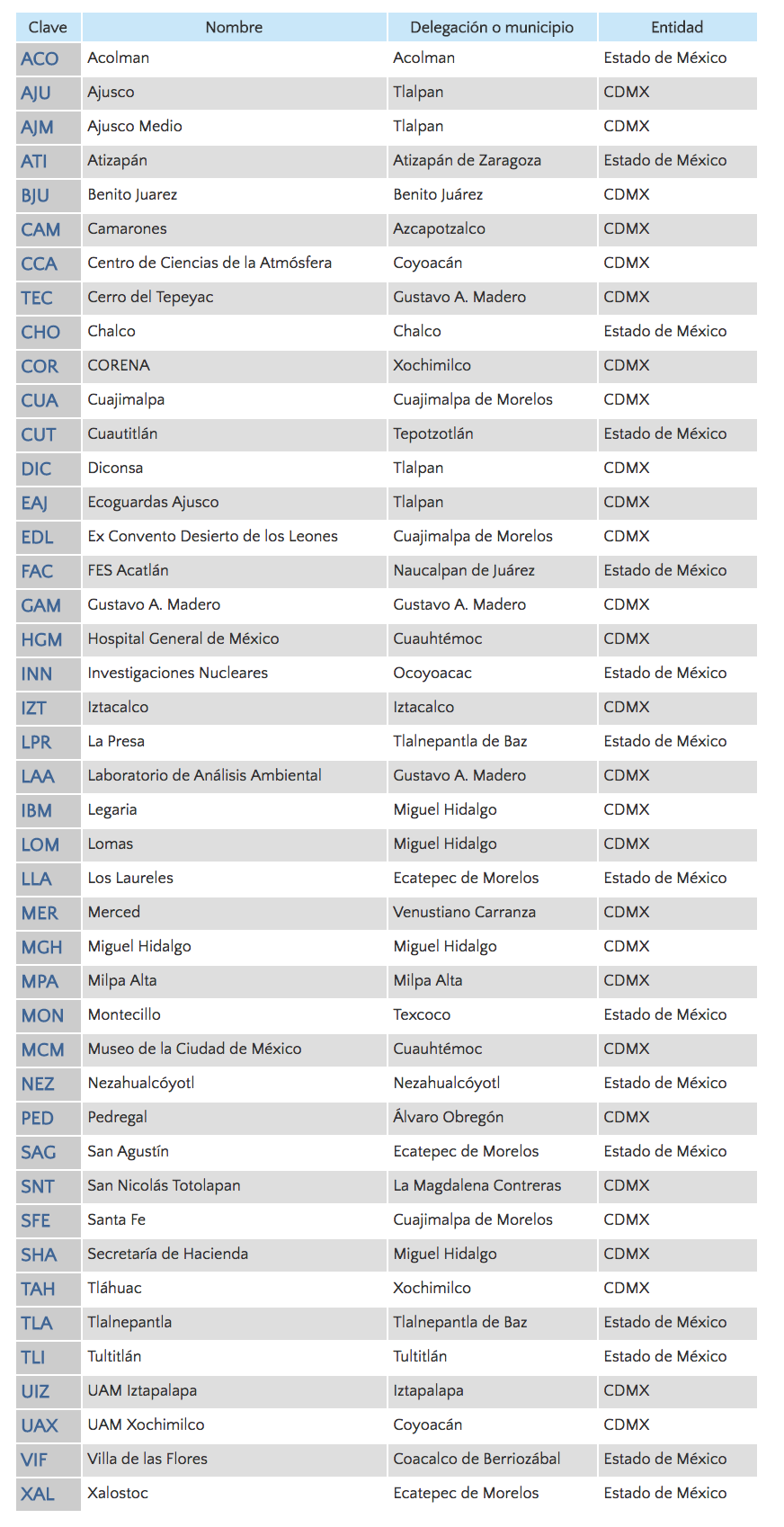
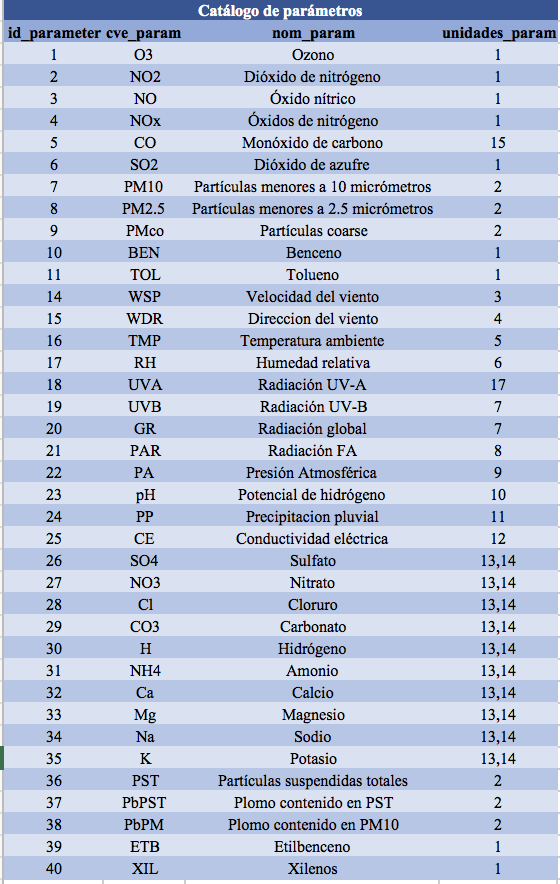


tabla 2. Catálogo de Parámetros Para Medir el Nivel de Contaminación Ambiental



1. **REFERENCIAS BIBLIOGRÁFICAS PREVIAS**

* SEDEMA
* SIMAT

1. **FUENTES SECUNDARIAS DE DATOS**

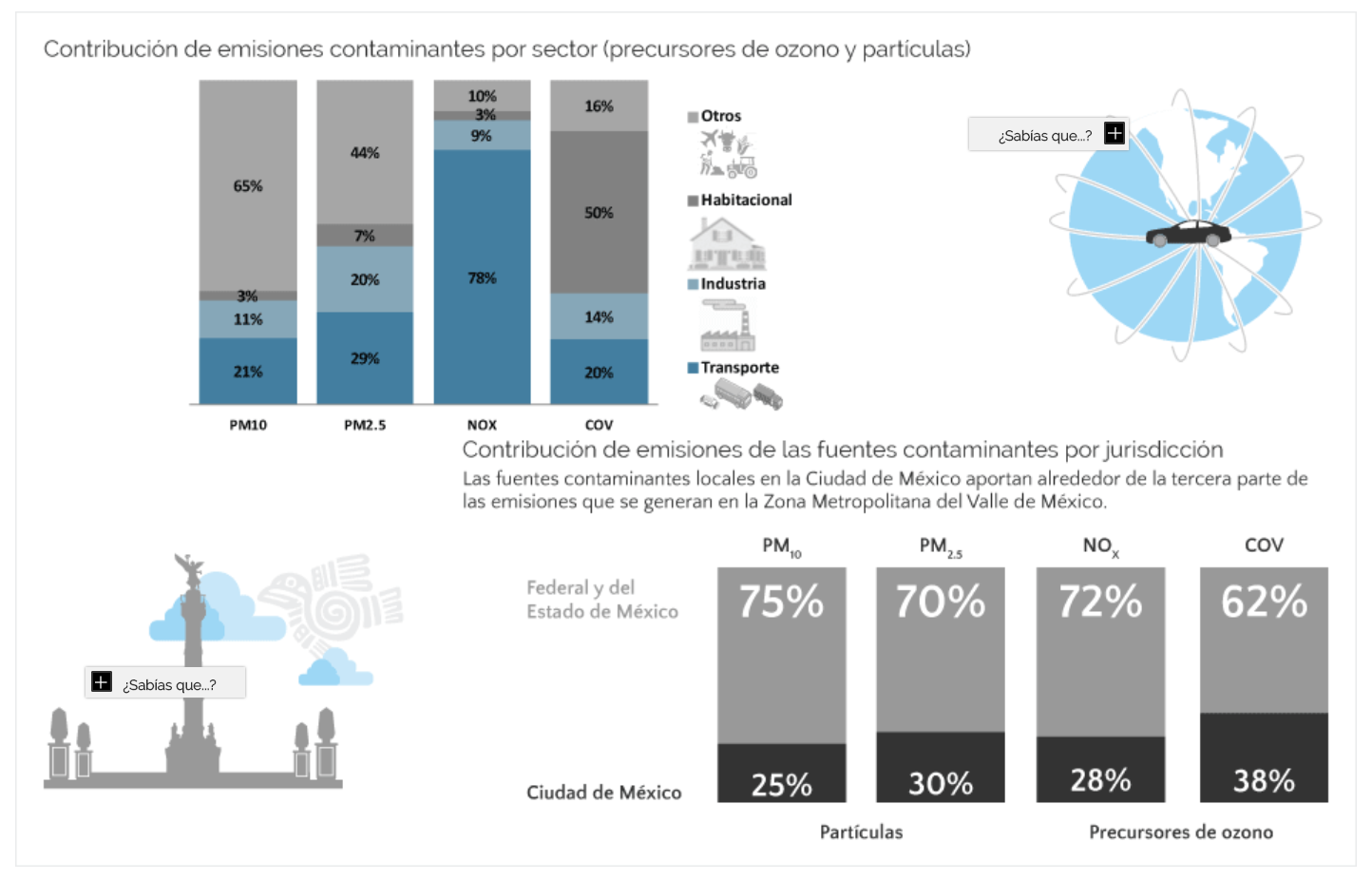
* STATISTA
* INEGI
* GOBIERNO CDMX
* GOBIERNO ESTADO DE MÉXICO

1. **TABLA DE CONTENIDO TENTATIVO DEL REPORTE FINAL**
2. **CRONOGRAMA**
3. **CURRICULA DE LOS CONSULTORES**

Vanessa Velázquez Mojica

* Estudiante Ing. en Negocios en el ITAM (2013-presente)

1. **ANEXOS Y APÉNDICES**

Anexo 1. Fuente: SEDEMA , 2017

1. (Marley, Gaffney, Ramos-Villegas, & Gonzalez, 2007) [↑](#footnote-ref-1)
2. STATISTA: (Climate Change; Expert(s) (Georgeson, et al. 2016), 2017) [↑](#footnote-ref-2)