**PROYECTO DE DATA SCIENCE**

**ANALISIS DEL DESEMPEÑO ACADEMICO DEL EXAMEN DE ESTADO DEL 2018 PARA EL INGRESO A LA EDUCACION SUPERIOR APLICANDO MINERIA DE DATOS**

**PRESENTADO POR:**

JESUS DAVID SUAREZ PEÑA

**PRESENTADO AL DOCENTE:**

ALVARO AGUSTIN OÑATE BOWEN

**UNIVERSIDAD POPULAR DEL CESAR**

**FACULTAD DE INGENIERÍA Y TECNOLOGÍAS**

**INGENIERIA DE SISTEMAS**

**VALLEDUPAR - CESAR**

**2020 - I**

**INTRODUCCIÓN**

El estudio de costos y tarifas de los servicios de acueducto y alcantarillado es la herramienta técnica y legal con la cual la empresa prestadora de servicio soporta la definición de los costos de administración, los costos de operación, costos de inversión y costos ambientales de tal forma que dichos costos convertidos en tarifas permiten recuperar los costos que generan a prestación de servicios garanticen la sostenibilidad financiera de la empresa siempre que sean coherentes con la capacidad económica tanto del municipio como de sus habitantes o usuarios

Las fórmulas tarifarias establecidas en la Resolución CRA 825de 2004, están estructuradas a través de componentes de acuerdo con la naturaleza de los costos en que incurren los prestadores de los servicios públicos domiciliarios de acueducto y alcantarillado. Los principales elementos en que se discriminan los costos dentro de la tarifa están relacionados con los procesos de administración, operación, inversión y los costos derivados de tasas ambientales.

Con base en la anterior resolución, las empresas de servicios públicos domiciliarios deben fijar los costos de referencia que se tomarán como base para la fijación de sus tarifas.

**ESTADO DEL ARTE**

Tras la expedición de la Constitución política de 1991 y la promulgación de la Ley 142de 1994, Colombia presentó avances en los indicadores de cobertura, calidad y continuidad para la prestación de los servicios públicos domiciliarios de acueducto y alcantarillado, especialmente en la zona urbana; no obstante, dichos avances no se dieron en igual proporción en el área rural, generando importantes brechas de cobertura entre áreas .Asimismo, aún persisten problemas en la calidad del agua suministrada a los usuarios en diferentes regiones del país, y en la continuidad en el servicio de acueducto. En este aparte se hace un análisis detallado del comportamiento de los principales indicadores sectoriales de los servicios, seguido por una breve caracterización del mercado y la revisión de algunos aspectos importantes asociados al esquema tarifario de los diferentes tipos de prestador en el mercado.

Para el caso del servicio público domiciliario de acueducto, en la información contenida en el CONPES 3918 -Estrategia para la implementación de los Objetivos de Desarrollo Sostenible (ODS) en Colombia, se muestra una cobertura para el año 2017 del 97,4% en suelo urbano, en cuanto al acceso a agua potable en suelo rural para ese año fue del 73,2%, para un total nacional del 92,4%.En cuanto al servicio público domiciliario de alcantarillado, para el mismo año, se tuvo una cobertura del 92,4% en suelo urbano, 70,1% en suelo rural y total nacional de 88,2%.

1. **COMPRENSION DEL NEGOCIO**

Es una empresa industrial y comercial colombiana de propiedad del municipio de [Medellín](https://es.wikipedia.org/wiki/Medell%C3%ADn_(Antioquia)), creada el 6 de agosto 1955. Es una empresa multilatina prestadora de servicios públicos (energía, gas, agua), la más grande de Colombia en este sector.

Como organización de servicios públicos domiciliarios tiene una historia para contar, con cifras y hechos de una responsabilidad social y ambiental que le da sentido a su origen, a su desarrollo y a su estrategia de negocios. Su objetivo principal es crear y preservar condiciones para asegurar la prestación de servicios sanitarios.

* **TARIFAS**

EPM, como empresa prestadora de los servicios públicos de acueducto, alcantarillado, energía y gas, no define sus tarifas de forma autónoma, es regulada por diferentes entidades del Estado como la Comisión de Regulación de Energía y Gas- [**CREG**](http://www.creg.gov.co/)  y la Comisión de Regulación de Agua Potable y Saneamiento Básico- [**CRA**](http://www.cra.gov.co/). Son estas quienes realizan la revisión de las tarifas y reglamentan sus variaciones.  
  
Dentro de la tarifa que pagan los clientes y usuarios de EPM en su factura, están incluidos componentes asociados a toda la operación de la infraestructura que se requiere para la prestación del servicio, así como las inversiones, el mantenimiento y los planes de expansión. También se contemplan todas las acciones necesarias para apalancar el crecimiento de los territorios donde la empresa ofrece sus servicios, así como el mantenimiento de los equipos.   
  
Al momento de revisar la factura de servicios públicos, más allá del valor total a pagar, es importante hacer un seguimiento a la cantidad de kilovatios (kW) de energía y a los metros cúbicos (m3) de agua y gas consumidos durante el mes. Del uso que se haga de estos servicios, dependerá el valor a pagar.  
  
Es importante tener en cuenta que los estratos socioeconómicos 5 y 6, hacen una contribución a los estratos 1, 2 y 3. El estrato 4 no subsidia ni es subsidiado. También que cada municipio cuenta con unos subsidios diferentes, lo que incide en que el costo de los servicios varíe de acuerdo con la localidad en la que se vive.

* 1. **DEFINICION DE LOS OBJETIVOS COMERCIALES**
* Brindar a los sectores información actualizada sobre las distintas tarifas por consumo.
* Proporcionar elementos al cliente para la realización de su autoevaluación de tarifas de consumos
* Identificar cuales son los servicios con mayores costos de tarifas
* Servir como fuentes de información para la construcción de indicadores de calidad de servicios
* Identificar el estrato mas afectados por las tarifas
* Monitorear la calidad de prestación del servicio que ofrece la empresa EPM a sus clientes.
* Ofrecer información que sirva como referente estratégico para el establecimiento de políticas empresariales.
  1. **VALORACION DE LA SITUACION**

Para la implementación del proyecto en donde se podrá visualizar información obtenida, aplicar análisis estadísticos, entre otros se tiene a disposición un hardware con pocos requisitos.

* **DATOS**: Se dispone de un repositorio que contiene toda La entidad tarifaria local aprueba las tarifas publicadas Facturación desde enero de 2017 hasta marzo de 2020. Los datos con lo que se llevara a cabo el desarrollo del proyecto se puede descargar desde el siguiente enlace: https://www.datos.gov.co/Funci-n-p-blica/Tarifas-Para-Servicios-De-Acueducto-y-Aguas-Residu/nfrm-mmfe
* **PERSONAL**: Si dispone de una persona capacitada para llevar a cabo el desarrollo del proyecto.
* **FACTOR DE RIESGO**: Este proyecto se deberá de completarse para mediados del mes de diciembre, por lo que el tiempo, en el que se llevara a cabo el proyecto es el principal factor de riesgo.

1. **RECURSOS DEL HARDWARE**

Se utilizará un hardware con pocas características ya que se aplicará la minería de datos con software sin alta necesidad de requerimientos, no obstante, deberá tener una capacidad media de almacenamiento debido a que los datos son de una magnitud media/baja.

1. **ORÍGENES DE DATOS Y ALMACENES DE CONOCIMIENTOS**

Los datos con los que se realizara el proyecto de descargaron de https://www.datos.gov.co. Esta página es utilizada como reportorio público donde podemos encontrar y descargar conjuntos de datos como son: datos del gobierno colombiano para investigar, desarrollar aplicaciones, crear visualizaciones e historias que pueden ser de utilidad para la realización de proyectos educativos o de cualquier tipo. Los datos se descargaron con extensión .csv (separados por comas) con 7.279 registros.

1. **OBJETIVOS DE LA MINERÍA DE DATOS**

El proyecto desarrollará principalmente capacidades predictivas, y también descriptivas, se tienen los siguientes objetivos:

* Preparar un conjunto de datos apropiados para aplicar las diferentes técnicas de la minería de datos.
* Proponer un modelo descriptivo para la visualización de los datos.
* Aplicar técnicas para predecir patrones ocultos.

1. **COMPRENSION DE LOS DATOS**
   1. **RECOPILACION DE DATOS INICIALES**
   2. **DATOS EXISTENTES**

El conjunto de datos sobre Tarifas Para Servicios De Acueducto y Aguas Residuales – EPMestá conformado por La entidad tarifaria local aprueba las tarifas publicadas. El dataset está conformado por 14 atributos o columnas en total, los cuales son

|  |  |
| --- | --- |
| **Atributo** | **Descripción** |
| **Municipio** | Municipio en el que se describirán las tarifas |
| **Sector** | Sector en el que se describirán las tarifas |
| **Estrato** | Estrato 1 Bajo - bajo Estrato 2 Bajo Estrato 3 Medio - bajo Estrato 4 Medio Estrato 5 Medio - alto Estrato 6 Alto |
| **Servicio** | Diferenciar si hace parte de Acueducto o Alcantarillado |
| **CargoFijo** |  |
| **CargoporConsumoMenor** |  |
| **CargoporConsumoMayor** |  |
| **CargoporConsumo** | Valor de la tarifa por consumo |
| **Suspensión** | Valor de la tarifa por suspensión |
| **Reinstalación** | Valor de la tarifa por reinstalación |
| **Reconexión** | Valor de a tarifa por reconexión del servicio |
| **Corte** | Valor de la tarifa de corte |
| **Año** | Año de la muestra |
| **Mes** | Mes de la muestra |