TRABAJO FINAL

**UNIVERSIDAD PERUANA**

**DE CIENCIAS APLICADAS**



**Trabajo Final**

**Profesor:** Cueva Chávez, Walter

**Integrantes:**

* Espiritu Cueva, Christian Joseph

**2020-02**

Tabla de contenido

[**1.** **Descripcion** 4](#_Toc52553607)

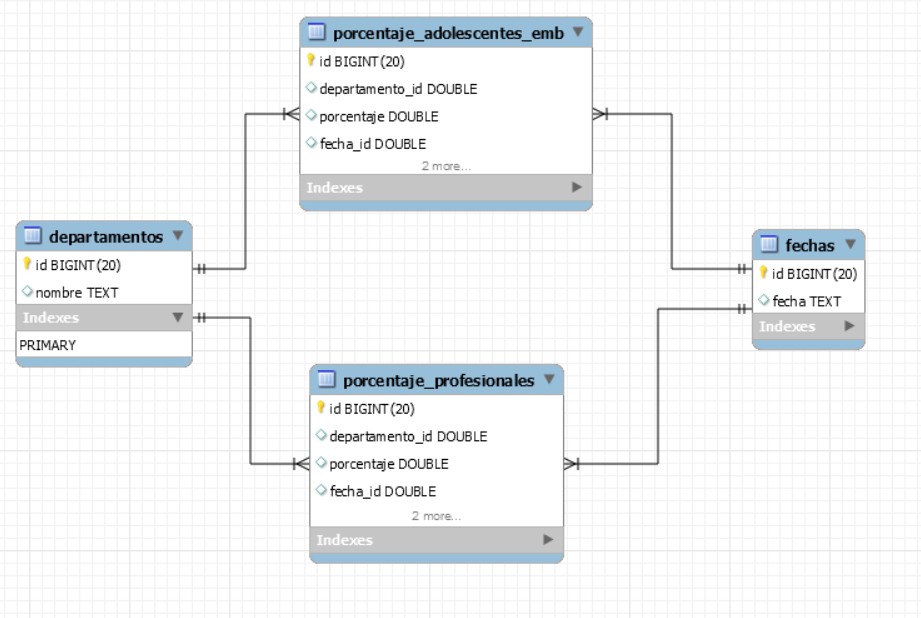
[**2.** **Procedimiento** 5](#_Toc52553608)

[**3.** **Conclusiones** 6](#_Toc52553609)

1. **Descripción**

El presente proyecto trata de dar una relación entre el porcentaje de adolescentes embarazadas por primera vez y profesionales de la salud que han atendido partos. Estos data sets, tienen fechas y regiones del Perú en común por lo que se vio viable su elección.

**Modelado de Base de Datos:**



Link de descarga de Script completo de base de datos:[**https://drive.google.com/file/d/1EGfJB-CHX8rXg\_uB4vbGATz42CacZyh0/view?usp=sharing**](https://drive.google.com/file/d/1EGfJB-CHX8rXg_uB4vbGATz42CacZyh0/view?usp=sharing)

1. **Procedimiento**

El proyecto inicia en primer lugar realizando la recolección de los datos, en este caso fueron dos archivos XLSX, uno de adolescentes embarazadas por primera vez y profesionales de la salud que han atendido partos a Nivel nacional en los años 2012 – 2018. En segundo lugar, se preparan los datos, optando por un muestreo de la población de datos, normalización, imputación eliminación de valores anómalos y outliers. En tercer lugar, se realiza el modelado de datos estructurados. En cuarto lugar, está la transformación y consultas exploratorias. En quinto lugar, se realiza la exploración visual de Datos. En sexto lugar, se implementan los modelos que en este caso se usaron KNN y regresión lineal. Finalmente, se realiza la exportación de documentos CSV y el envió de correos.

LINK DEL REPOSITORIO: <https://github.com/NightBright/TrabajoFinalRShiny>

3. Conclusión

En síntesis, los datos obtenidos de los archivos XLSX, que contienen información sobre el porcentaje de adolescente que son madres primerizas y los profesionales que han atendido embarazos entre los años 2011 a 2018. Primero se realizó la preparación de datos, el cual se editan los datos para que puedan ser ejecuta adecuadamente sin valores anómalos ni outliers. Segundo, el modelado de datos estructurados, es aquí donde se suben todos los datos a la base de datos desplegada en AWS. Tercero, transformación y consultas exploratorias, es donde se realizan las consultas llamando a la base de datos y operaciones de estadística descriptiva. Cuarto, exploración visual de datos, en el cual se realizaron gráficos estadísticos. Quinto, modelo, es donde se realizó la regresión lineal del porcentaje de adolescente embarazadas por primera vez junto con porcentaje de profesionales que atendieron partos. Sexto, la exploración y comunicación, es donde se realiza la exportación de datos en formato CSV. Finalmente, la implementación adicional en el cual se realizó el envió de correos mediante RStudio.