

PRACTICA 1

1.- Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.

El contexto de la investigación, es poder obtener de la página publica de consultas vehicular a través del número de placa los modelos de carros, tipo de servicio que prestan, años de los vehículos, etc. Para realizar análisis de posibles provincias donde se podrían colocar ciertos tipos de vehículos, podría ser de acuerdo a su zona geográfica, modelo más utilizado, verificar la vida útil, etc.

Sitio web escogido: https://consultas.atm.gob.ec/PortalWEB/paginas/clientes/clp_criterio_consulta.jsp

2.- Definir un título para el dataset. Elegir un título que sea descriptivo.

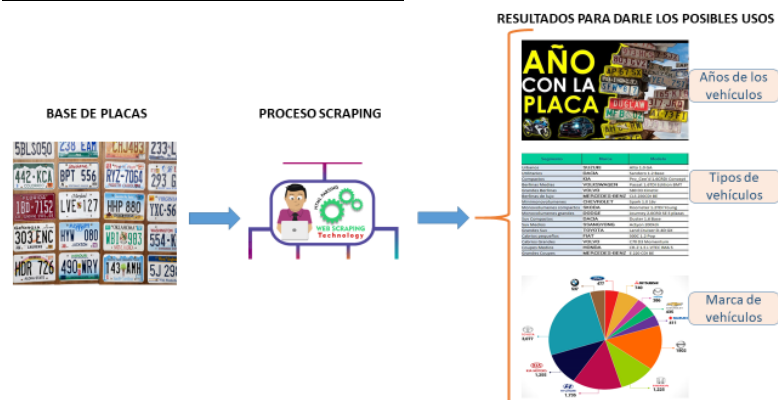
Descripción del Modelo de vehículo a través de su placa

3.- Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).

El dataset obtenido con el web scraping, se trata de una muestra de datos recolectados a través de las placas de vehículos, esto con el fin de obtener qué modelo es, año del vehículo, marca del vehículo, color del vehículo, provincia (de que parte del país es), pago de matrícula, con el fin de proponer modelos de vehículos a vender.

4.- Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente

Proceso de extracción del dataset



Detalle del dataset: muestra 937 registros obtenidos

TIPO DE SERVICIO	REGISTROS	TIPO DE VEHICULO	REGISTROS	MARCA VEHICULO	REGISTROS
USO PARTICULAR	766	AUTOMOVIL	258	CHEVROLET	261
USO PUBLICO	128	CAMIONETA	216	KIA	79
USO ESTATAL	41	MOTOCICLETA	199	HYUNDAI	68
CUERPO CONSULAR	2	VEH. UTILITARIO	115	MAZDA	52
TOTAL	937	CAMION	67	SUZUKI	49
		SIN VALOR	30	TOYOTA	45
		VEHICULO ESPECIAL	22	NISSAN	44
		AUTOBUS	22	HINO	38
		UNIDAD DE CARGA Y REMOLQUE	8	MOTOR UNO	30
		TOTAL	937	BAJAJ	28
				LONCIN	19
				RENAULT	19
				GREAT WALL	15
				RANGER	14
				VOLKSWAGEN	13
				FORD	12
				OTROS	151
				TOTAL	937

5.- Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

CAMPO	DETALLE
NUMERO_PLACA	SE REGISTRA EL NUMERO DE PLACA DEL VEHICULO
MARCA	SE REGISTRA LA MARCA DEL VEHICULO
COLOR	SE VERIFICA EL COLOR DEL VEHICULO
ANIO DE MATRICULA	SE DETERMINA EL ULTIMO AÑO DE MATRICULACION
MODELO	DETALLE DEL MODELO DE VEHICULO
CLASE	TIPO DE VEHICULO
FECHA DE MATRICULA	FECHA DE MATRICULACION
ANIO	AÑO DEL VEHICULO
SERVICIO	TIPO DE SERVICIO PARA EL CUAL SE LO UTILIZA
FECHA DE CADUCIDAD	FECHA DE CADUCIDAD DE LA MATRICULA

Periodo del tiempo: Vehículos modelos desde el 2008 hasta el 2019

Se los recogió ingresando una lista de placas almacenadas en una base para obtener los campos descritos en la figura anterior por medio del web scraping.

6.- Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).

Agradecimientos a la ATM, Agencia municipal de tránsito que es la entidad que permite esta consulta por usuario ingresando a su página web.

Investigaciones anteriores: no he visto que utilicen scraping para realizar análisis de posibles ventas por provincias o por marcas de carro en mi ciudad en base a la página de la ATM.

7.- Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.

Este dataset nos podría ayudar a ver qué provincia tiene más vehículos por tipo, modelo, por marca, de esta manera y más otros análisis más podríamos graficar el país y determinar a qué provincia se le debería promocionar más un determinado tipo de vehículo.

Ejemplo sencillo del estudio:



8.- Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección

Released Under CC BY-NC-SA 4.0 License:

Nos permite compartir material en cualquier formato, adicional nos permite editar el material para uso de nuestros análisis.

- La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

9.- Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

Se adjunta código en Github

Pego el código:

proceso que realiza la extracción de datos aplicando la técnica del web scraping y poder generar el dataset para el análisis.

```
def consultaPlaca(placa):
    data_vehicle = []
    data_vehicle.append(placa)

    try:
        quote_page =
'http://consultas.atm.gob.ec/PortalWEB/paginas/clientes/clp_grid_citaciones.jsp?ps_tipo_identificacion
=PLA&ps_identificacion=' + placa + '&ps_placa='
        page = requests.get(quote_page)
        soup = BeautifulSoup(page.text, 'html.parser')
        table = soup.find('table', {'cellpadding': "2"})
        rows = table.find_all('tr')
        for tr in rows:
            # print(tr)
            for wrapper in tr.find_all(lambda tag: tag.name == 'td' and tag.get('class') ==
['detalle_formulario'] and tag.get('class') == ['detalle_formulario']):
                data_vehicle.append(wrapper.text)
                # print(wrapper.text)
                # write_csv(wrapper.text)
            # print(data_vehicle)
        return data_vehicle
    except Exception as e:
        print('Datos no disponibles')
        print(e)
        return data_vehicle
```

10.- Dataset. Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

<https://zenodo.org/record/4263471#.X6iawGhKjIU>

Datos de los autores:

CONTRIBUCIONES	FIRMA
INVESTIGACION PREVIA	ELOY BRAVO
REDACCION DE LAS RESPUESTAS	ELOY BRAVO
DESARROLLO DEL CODIGO	ELOY BRAVO