

Identificación de Tendencias de Crecimiento y Clasificación de Municipios por Saturación de Vehículos

PRIMERA ENTREGA: PREPARACIÓN DE DATOS PARA SU MODELADO

Aura Melina Gutierrez Jimenez

Paola Covarrubias Viveros

Bernardo Orozco Quirarte

Carlos Chihenseck Blanco

MINERIA DE DATOS | DIANA GABRIELA GOMEZ MARTINEZ

Contenido

Entendimiento del negocio	2
Objetivo del negocio	2
Objetivo de minería de datos	2
Entendimiento de los datos	2
Origen de los datos/ Dataset	2
Inconsistencias	3
Preparación de los datos	3
Selección y limpieza	3
Integración de datos	4
Transformación	5
Visualización.....	5

Entendimiento del negocio

Objetivo del negocio

El crecimiento vehicular en los municipios de Jalisco a lo largo de 41 años es un indicador clave para evaluar la capacidad de la infraestructura vial y la efectividad de las políticas de movilidad. El objetivo principal **es identificar los municipios que están experimentando el mayor crecimiento en el número de vehículos y evaluar su riesgo de saturación vial**. Este análisis permitirá a los responsables de la planificación urbana y de la infraestructura diseñar estrategias para evitar congestiones de tráfico y priorizar inversiones en infraestructura vial.

Objetivo de minería de datos

El análisis del crecimiento vehicular en los 125 municipios de Jalisco, considerando 41 años de datos (1980-2020), permitirá identificar patrones y tendencias en el crecimiento del parque vehicular. Utilizando técnicas de minería de datos como **clustering y análisis de series temporales, se podrán clasificar los municipios según su riesgo de saturación vehicular**. Esta clasificación servirá para diseñar estrategias específicas de desarrollo vial y movilidad urbana, adaptadas a las características de cada grupo de municipios. Por otro lado, podríamos utilizar técnicas de análisis exploratorio y algoritmos de minería de datos, **como análisis de tendencias y modelos predictivos para identificar los municipios con mayor crecimiento en el número de vehículos**.

Entendimiento de los datos

Origen de los datos/ Dataset

El dataset utilizado proviene del **Instituto de Información Estadística y Geográfica del Estado de Jalisco (IIEG)** y contiene información sobre los vehículos de motor registrados en circulación por municipio en el estado de Jalisco durante el período comprendido entre el 1 de enero de 1980 y el 31 de diciembre de 2020. Los datos han sido recopilados y publicados por la Dirección de Información Estadística de Gobierno, Seguridad Pública e Impartición de Justicia del estado de Jalisco. El formato en el que los datos fueron obtenidos fue en xlxs (Excel), pero debido al formato, decidimos cambiar el archivo a csv.

El dataset contiene información detallada de 125 municipios del estado de Jalisco, además del estado en su conjunto. Cada fila representa un municipio o el estado en su conjunto, y las columnas contienen los valores anuales de vehículos registrados. Los

datos de crecimiento vehicular a lo largo de los años permiten identificar patrones de crecimiento en diferentes municipios. Esto ayudará a determinar cuáles municipios están experimentando un crecimiento acelerado en su parque vehicular y requieren atención prioritaria.

Cantidad de registros:

- 125 municipios + 1 registro para el total del estado = 126 registros en total.
- Cada registro tiene 41 columnas de años (1980 a 2020) más una columna para el nombre del municipio o estado.
- Registros totales = 5,166 registros
- Registros totales después de limpieza = 2,875

Relación con el problema de negocio: Los datos de crecimiento vehicular permiten identificar cuáles municipios están en riesgo de saturación vial, información clave para la planificación de infraestructura y políticas públicas.

Inconsistencias

Existen registros en blanco, por lo que hay municipios con datos incompletos. Estos datos faltantes pueden afectar la precisión de los análisis.

Preparación de los datos

Selección y limpieza

Eliminar columnas irrelevantes o con demasiados valores faltantes, como municipios con datos inexistentes para la mayoría de los años. Decidimos eliminar los registros de todos los municipios en el rango 1980 a 1996, debido a la inconsistencia de los datos y la cantidad exagerada de registros nulos. El mantener esa cantidad de nulos en nuestro dataset podría perjudicar nuestros modelos y generar valores erróneos.

1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
487,847	506,434	539,891	555,995	586,390	627,735	674,354	741,510	789,104	1,006,337	1,061,022	1,005,306	1,016,441	1,077,416	1,039,757	1,114,421	1,095,109	1,174,637	1,441,016	1,114,116	1,540,129	2,046,373
1,06	1,100	1,154	1,185	1,199	1,163	2,038	2,273	2,412	2,539	2,727	2,816	3,014	3,171	2,701	2,839	3,119	4,440	4,330	4,402	5,007	5,502
1,132	1,402	1,460	1,463	1,365	1,463	1,571	1,723	1,851	2,375	2,314	2,343	2,375	2,629	2,441	2,639	3,096	3,414	4,052	3,702	3,979	4,343
-	-	-	-	-	-	-	-	-	3,225	3,363	3,208	3,330	3,414	3,502	3,300	3,240	3,626	4,364	4,065	4,963	5,267
-	-	-	-	-	-	-	-	-	-	-	-	-	789	954	939	1,174	1,468	1,677	1,709	1,647	1,928
2,044	2,360	3,107	3,209	3,363	3,540	4,474	4,300	5,239	10,464	8,807	9,200	9,395	10,359	10,423	9,329	10,375	12,027	12,029	13,305	15,302	15,308
-	-	-	-	-	-	-	-	-	1,029	1,100	1,106	1,091	1,097	952	974	1,236	1,295	1,643	1,961	3,022	1,905
5,04	4,320	4,349	5,261	5,606	6,052	7,230	8,009	8,582	10,493	10,680	14,139	15,353	15,272	15,474	20,043	19,148	20,269	23,062	24,033	26,245	27,363
-	-	-	-	-	-	-	-	-	1,857	2,060	2,026	2,053	2,078	2,103	2,296	2,350	2,861	2,855	3,073	3,176	3,405
-	-	-	-	-	-	-	-	-	-	-	-	-	647	634	704	765	855	977	1,007	1,047	906
-	-	-	-	-	-	-	-	-	-	-	-	-	632	601	674	762	830	864	890	951	1,010
-	-	-	-	-	-	-	-	-	-	-	-	-	737	744	716	901	1,080	1,056	1,254	1,282	1,391
2,102	2,307	3,091	3,395	3,143	3,302	4,835	5,090	5,467	6,527	6,960	6,829	7,062	8,246	8,607	8,053	9,472	11,321	14,009	14,044	16,409	14,419
-	-	-	-	-	-	-	-	-	-	-	-	-	950	1,000	1,040	1,038	1,087	1,339	1,335	1,500	1,077
4,103	4,316	4,737	4,302	5,270	5,472	6,381	6,778	7,300	9,314	8,601	10,320	10,706	11,917	9,305	9,479	10,649	11,420	11,862	12,409	13,830	15,432
-	-	-	-	-	-	-	-	-	3,201	3,376	3,280	3,341	3,638	3,360	3,371	4,236	5,409	6,051	6,429	6,193	6,739
-	-	-	-	-	-	-	-	-	1,850	1,743	1,889	1,669	1,703	1,721	1,780	2,039	2,316	2,779	2,889	3,300	3,315
3,332	4,264	4,473	4,717	5,076	5,461	5,787	6,358	6,810	6,636	7,324	6,302	7,435	7,200	7,337	7,695	8,642	9,393	10,323	10,809	12,192	13,519
-	-	-	-	-	-	-	-	-	-	-	-	-	352	351	353	366	411	487	487	479	581
-	-	-	-	-	-	-	-	-	-	-	-	-	946	940	966	1,044	1,174	1,279	1,370	1,555	1,574
1,615	1,591	1,602	1,723	1,648	1,540	2,240	2,244	2,414	2,849	3,003	2,810	2,931	2,272	2,074	2,091	3,301	3,301	4,147	4,105	4,477	4,391
2,096	2,843	2,366	2,577	3,002	3,029	3,390	3,720	4,095	4,626	4,203	5,002	5,188	5,420	4,104	4,403	5,290	5,004	5,232	6,044	6,427	6,782
6,407	6,364	5,388	10,305	11,329	11,841	13,206	14,465	15,663	17,507	16,795	19,027	17,464	14,234	16,517	16,031	18,436	19,352	21,053	23,331	25,439	26,398
1,238	1,235	1,506	1,649	1,746	1,940	1,249	2,240	2,238	3,462	3,328	3,453	3,096	3,306	4,092	3,373	3,347	4,438	5,171	5,097	6,230	6,534
-	-	-	-	-	-	-	-	-	2,653	2,700	2,803	2,765	2,940	2,639	2,708	3,098	3,302	3,526	3,690	4,207	4,300
-	-	-	-	-	-	-	-	-	-	-	-	-	901	1,051	1,032	1,273	1,532	1,516	1,451	1,967	1,607
-	-	-	-	-	-	-	-	-	-	-	-	-	763	781	779	936	1,254	1,390	1,662	1,641	1,890
-	-	-	-	-	-	-	-	-	-	-	-	-	716	734	764	903	930	970	930	1,044	1,032
-	-	-	-	-	-	-	-	-	2,303	2,534	2,704	2,484	2,397	2,080	2,191	3,337	3,363	4,661	4,303	5,044	5,145
1,644	2,053	2,343	2,558	2,717	2,914	4,711	5,170	5,497	6,160	5,540	6,081	7,318	7,487	5,846	6,164	5,839	7,248	8,433	8,045	9,303	9,724
-	-	-	-	-	-	-	-	-	-	-	-	-	211	216	220	211	235	210	281	291	316
-	-	-	-	-	-	-	-	-	-	-	-	-	487	505	520	629	714	739	1,004	1,052	1,302
-	-	-	-	-	-	-	-	-	3,654	4,322	4,039	4,303	4,453	4,647	4,735	5,361	6,053	6,393	6,769	7,192	7,509
-	-	-	-	-	-	-	-	-	-	-	-	-	716	727	427	432	419	432	616	616	616
1,094	1,301	2,004	2,550	2,912	3,086	3,435	3,760	4,060	5,291	5,214	6,250	6,007	5,205	6,206	6,890	7,896	9,305	8,332	9,817	10,901	10,093
1,510	1,721	1,907	2,168	2,407	2,506	2,787	3,072	3,289	3,620	3,833	3,155	4,093	4,104	4,345	4,303	4,632	5,156	6,257	6,294	6,693	7,091
-	-	-	-	-	-	-	-	-	-	-	-	-	432	437	533	640	841	1,006	1,075	1,346	1,202
336,905	393,225	347,414	379,369	397,051	425,570	436,995	460,251	508,722	634,293	689,207	620,146	593,272	400,359	470,436	416,053	475,897	405,430	593,323	627,075	677,238	688,044

Imputación de valores faltantes: Utilizamos el método de cuartil inferior para completar los valores faltantes en los registros donde solo falten algunos años, ya que utilizar la media no nos pareció lo adecuado debido a las diferencias entre los volúmenes de vehículos registrados entre municipios.

Conteo de valores nulos después de eliminar las co		Estado o Municipio	
Estado o Municipio	0	1997	0
1997	5	1998	0
1998	4	1999	0
1999	4	2000	0
2000	4	2001	0
2001	4	2002	0
2002	4	2003	0
2003	4	2004	0
2004	4	2005	0
2005	4	2006	0
2006	4	2007	0
2007	4	2008	0
2008	4	2009	0
2009	4	2010	0
2010	3	2011	0
2011	3	2012	0
----	-		

Integración de datos

En nuestro caso, solamente utilizamos un dataframe principal, en el cual operamos y realizamos funciones de agrupación (groupby) y de normalización con la librería scikit-learn, para facilitar la lectura de la densidad de los valores de los registros. Al aplicar la función de agrupación, logramos consolidar los datos por municipio, obteniendo así una visión más clara del total de vehículos registrados.

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	...	2013	2014	2015	2016	2017	2018	2019	2020	Total Vehiculos
Estado o Municipio																				
Acatitlán	2717	2751	2896	3795	4446	4708	4882	5083	5622	6145	...	9365	9929	10580	10957	11386	11792	12411	12874	180661
Acatlán de Juárez	2629	2645	2699	3056	3614	4052	3782	3979	4363	4959	...	6160	6432	6668	6960	7261	8125	8590	9067	123069
Ahualulco de Mercado	3414	3502	2900	3240	3626	4568	4885	4850	5267	5763	...	7990	8259	8528	8974	9610	10264	10723	11296	161115
Amacueca	799	954	999	1174	1468	1677	1709	1867	1928	2055	...	2559	2603	2661	2720	2833	3147	3177	3232	51954
Amatitán	1749	1762	1889	2391	2733	3305	3317	3525	3328	3318	...	4502	4788	4974	5228	5735	6282	6798	7286	96254

5 rows × 26 columns

Transformación

Posteriormente, implementamos la normalización para ajustar los valores de la nueva columna 'Total Vehículos', lo que nos permitió comparar fácilmente las cifras entre los diferentes municipios y observar tendencias en la densidad de vehículos registrados. Esta metodología nos permitió simplificar el análisis y mejorar la interpretabilidad de los resultados.

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	...	2013	2014	2015	2016	2017	2018	2019	2020	Total Vehículos	Total Vehículos Normalized
Estado o Municipio																					
Acatic	2717	2751	2896	3795	4446	4708	4882	5083	5622	6145	...	9365	9929	10580	10957	11386	11792	12411	12874	180661	0.009616
Acatlán de Juárez	2629	2645	2699	3056	3614	4052	3782	3979	4363	4959	...	6160	6432	6668	6960	7261	8125	8590	9067	123069	0.006387
Ahualulco de Mercado	3414	3502	2900	3240	3626	4568	4885	4850	5267	5763	...	7990	8259	8528	8974	9610	10264	10723	11296	161115	0.008520
Amacueca	799	954	999	1174	1468	1677	1709	1867	1928	2055	...	2559	2603	2661	2720	2833	3147	3177	3232	51954	0.002401
Amatitán	1749	1762	1889	2391	2733	3305	3317	3525	3328	3318	...	4502	4788	4974	5228	5735	6282	6798	7286	96254	0.004884

5 rows × 26 columns

Visualización

Durante el proceso hacer uso de streamlit para visualizar los resultados, por ejemplo, o Antes y después de la limpieza: Mostrar gráficos que comparen los datos antes y después de la eliminación de columnas irrelevantes o la imputación de valores faltantes. Estas visualizaciones ayudarán a documentar claramente las decisiones tomadas en cada paso, mostrando visualmente los efectos de las modificaciones aplicadas a los datos.

Estado o Municipio	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Asoyac	1,040	1,040	1,240	1,201	1,330	1,330	1,300	1,311	1,120	1,040
Autlán de Navarro	9,105	9,479	10,669	11,425	11,862	12,409	13,830	15,622	17,660	19,498
Ayotlán	3,180	3,357	4,292	5,409	6,051	6,435	6,193	6,739	7,123	7,686
Ayutla	1,721	1,780	2,099	2,516	2,779	2,889	3,100	3,315	3,652	3,943
Bolaños	331	333	388	411	437	467	479	528	587	645
Cabo Corrientes	948	966	1,044	1,174	1,279	1,370	1,258	1,374	1,576	1,815
Casimiro Castillo	2,894	2,851	3,101	3,801	4,147	4,185	4,477	4,811	5,319	5,850
Cañadas de Obregón	941	936	1,154	1,321	1,441	1,498	1,470	1,499	1,660	1,805
Chapala	5,846	6,154	5,899	7,348	8,453	8,845	9,930	9,724	11,674	13,250
Chimaltitán	216	220	217	235	270	281	291	316	352	377
Chiquilistlán	505	520	629	574	739	1,024	1,052	1,122	1,172	1,258

Filtrado de datos por municipio (datos procesados):

Selecciona un municipio

Acatic													
	Estado o Municipio	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	
1	Acatic	2,717	2,751	2,896	3,795	4,446	4,708	4,882	5,083	5,622	6,145	6,145	

Visualización de los datos de Acatic

