**ESTUDIO DE LOS FACTORES DE INCIDENCIA EN SINIESTROS VIALES EN EL DEPARTAMENTO PARANÁ, REGISTRADOS POR LA DIVISIÓN DE ACCIDENTOLOGÍA VIAL DE LA POLICÍA DE ENTRE RÍOS**

**Alejandro Rouiller – Zacarías Ojeda – Melisa Fernández**

**Resumen:**

Según los datos de la Organización Mundial de la Salud (OMS) los siniestros de tránsito provocan cada año entre 20 y 50 millones de heridos y un total de 1,27 millones de muertes, y son considerados como uno de los grandes problemas de salud pública a nivel mundial, ya que su impacto no solo se limita a pérdidas materiales, sino lo más importante, vidas humanas y las consecuencias que lo preceden para un núcleo familiar con afectaciones psicológicas, físicas y económicas, que también se manifiestan en los altos costos para los servicios de salud. (OMS, 2017)

La Argentina es uno de los países con más alto índice de mortalidad por este motivo, y pese al compromiso asumido con la OMS en 2011, no se ha logrado reducir en los últimos años la cantidad de fatalidades por siniestros viales, ocupando la provincia de Entre Ríos la séptima posición en la escala nacional.

Los siniestros viales vienen definidos a través de una serie de variables explicativas, que implican la concurrencia de diversos factores, referentes tanto al vehículo, al entorno, a las características de la vía y como al propio factor humano. La recopilación de estos datos, permite consultar una gran cantidad de estadísticas, pero el valor potencial en las mismas reside en el conocimiento oculto que pueda extraerse de los datos (Lopez et al., 2012). La Minería de Datos provee de las técnicas que facilitan el estudio y procesamiento avanzado de los datos permitiendo encontrar patrones, predecir comportamientos, y describir asociaciones entre variables que resultan imperceptibles sin este análisis.

El principal objetivo de esta investigación es analizar los siniestros viales registrados en el departamento de Paraná durante el primer trimestre de 2018 por la División de Accidentología de la Dirección Criminalística de la Policía de Entre Ríos con el objetivo de determinar asociaciones entre variables para predecir la severidad de los mismos.

Para esta investigación, la técnica de Árboles de Decisión (ADD) es una de las que resulta más apropiada, ya que no suele suponer ninguna relación previa entre las variables. Además, permite la extracción de determinadas Reglas de Decisión (RD) del tipo “SÍ-ENTONCES” (Kashani et al., 2011), que pueden ser usadas para la comprensión del suceso, la probabilidad de que suceda y la identificación de las principales variables que determinan su gravedad.

Para la exploración de los datos se realizó pruebas Chi cuadrado, atendiendo encontrar dependencias entre cada variable y el tipo de siniestro, y un Análisis de Correspondencia Múltiple para ver la asociación entre las variables y categorías de estudio.

Para el análisis de la geolocalización se realizó un Escalonamiento Multidimensional, y luego se analizaron conglomerados en el mapa de Paraná respecto a los factores de dependencia.

Se espera que los resultados obtenidos aporten información que contribuya con la División de Criminalística de la Policía de Entre Ríos y a las políticas de prevención que delinean sus especialistas, para reducir el impacto socio-económico causado por los accidentes de tráfico, mejorar la seguridad vial de la ciudad y que sean el puntapié inicial para una investigación de mayor envergadura.

**Situación problemática y planteo del problema**

En el año 2011, la Agencia Nacional de Seguridad Vial adhirió al Decenio para la Acción para la Seguridad Vial 2011-2020 (disposición 92/2011), medida promovida por la Organización Mundial de la Salud y dirigida a estabilizar y reducir la cifra mundial de víctimas fatales en siniestros de tránsito prevista para el año 2020 (Observatorio de Seguridad Ciudadana de Avellaneda, Otamendi & Rey, 2015).

Sin embargo, las cifras de muertes por siniestros viales en el país, y en Entre Ríos siguen sin disminuir y resultan alarmantes. Se producen por este motivo en la provincia alrededor de 190 muertes al año, que derivan en consecuencias psicológicas (traumas y dificultades para la vida cotidiana) y físicas (discapacidades permanentes, por ejemplo), además de consecuencias económicas, aumentando el gasto tanto a nivel público (costos de traslados, primeros auxilios, rehabilitación, etc.) como a nivel administrativo (costos policiales, judiciales, etc.), entre otras.

La División de Accidentología de la Dirección Criminalística de la Policía de Entre Ríos Delegación Paraná lleva una ardua tarea al tener que intervenir en más de 800 siniestros anuales, si bien su tarea es la de asistir en caso de que los siniestros sean graves, muchos de estos casos aparentan mayor gravedad de la que en realidad revisten, generando una sobrecarga en su tarea diaria. Los agentes de esta división realizan un registro y seguimiento de cada caso donde actúan, de manera que es posible determinar la gravedad de los siniestros ocurridos en la ciudad, y los diversos factores que incidieron en los mismos.

Esta investigación a través de técnicas de Minería de Datos se plantea asociaciones entre los factores que intervienen en siniestros viales, con el objeto de descubrir patrones que ayuden a la prevención y a la disminución de las secuelas que se producen como resultado de los mismos.

**Objetivos de la investigación.**

* Determinar asociaciones, y relaciones entre las variables que intervienen en los siniestros viales.

**Antecedentes**

En Argentina, se encontró un solo estudio en referencia con la temática, denominado “Análisis de variables asociadas en siniestros viales dentro de un área georreferenciada para establecer probabilidades de hospitalización. Caso de aplicación”. Sus autores Martínez Micakoski , F., y River, J. J., identificaron mediante un análisis de asociación de variables, y particionado la región de estudio en cuadrículas, los factores Humano/Vehículo/Ambiente que intervienen en la hospitalización de un siniestrado dentro de las subparticiones de la región de Trenque Lauquen.

Fuera del país es posible encontrar diversas investigaciones sobre la temática:

En Canadá en el año 2002, Bedard et al. , aplicando regresión logística encontraron que la prevención de fatalidades radicaba en el mayor uso del cinturón de seguridad, en la reducción de velocidad e impactos del lado del conductor.

En China, Kong y Jang en 2010, estudiaron los siniestros de peatones con el objeto de investigar la relación entre la velocidad de impacto y el riesgo de los peatones en las colisiones de vehículos. Para el análisis desarrollaron modelos de regresión logística de los riesgos de muertes y lesiones en los peatones, en términos de velocidad de impacto del vehículo. Encontraron que el riesgo de mortalidad de peatones es de 26% a 50 km/h, 50% a los 58 km/h, y el 82% a los 70 km/h.

En España en el 2011, De Oña, J., Oqab, R. y Calvo, F., presentaron un estudio titulado *“Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks”* (de Oña, Oqab Mujalli, & Calvo, 2011). Donde utilizaron redes Bayesianas (BNs) para clasificar los siniestros viales en función a la gravedad de la lesión. En el estudio analizaron 1536 siniestros ocurridos en las carreteras rurales de España, y 18 variables que representan los factores que contribuyen a la ocurrencia del accidente. Encontraron que las variables que mejor identifican los factores que se asocian con un accidente de muerte o lesiones graves es el tipo de accidente, la edad del conductor, la iluminación y el número de lesiones.

En el mismo país, López Maldonado, G. en 2013 presentó una tesis doctoral titulada “Análisis de la Severidad de Siniestros de Tráfico, utilizando Técnicas de Minería de Datos”; en donde propone un estudio en profundidad de la accidentabilidad de las carreteras de Granada, España. La investigación utilizó árboles de decisión para extraer patrones de siniestros cuya severidad es grave usando reglas de decisión. Obteniendo una precisión del 55.57 % en la clasificación de la severidad de los siniestros.

En Chile, Rubio et al. en 2013, desarrolló y evaluó modelos de predicción que estiman el número de personas lesionadas y fallecidas en siniestros viales, relacionado con las causas que produce el accidente. Utilizó para ello metaheuristicas y redes neuronales artificiales en combinación con algoritmos de optimización.

**Referentes Conceptuales**

En el siguiente trabajo se entenderá por siniestros viales a cualquier hecho de tránsito con implicación de al menos un vehículo en movimiento, que tenga lugar en una vía pública o en una vía privada a la que la población tenga derecho de acceso, y que tenga como consecuencia al menos una persona herida o muerta.

Para categorizar a los siniestros viales se tiene en cuenta la gravedad de las víctimas involucradas, tomando como criterio a la víctima de mayor gravedad como definitoria del tipo de siniestro, categorizando en:

Siniestros leves: No hubo ninguna persona involucrada en el siniestro que fuera un herido grave.

Siniestros graves: Al menos una persona en el siniestro resulto ser un herido grave.

Se entiende por herido grave a cualquier persona herida que ha sido hospitalizada durante más de 24 horas, se incluyen en esta categoría personas fallecidas por causas del siniestro.

**Técnicas de análisis de información**.

La Minería de Datos integra numerosas técnicas de análisis de datos y extracción de modelos, para encontrar patrones, describir tendencias y regularidades, predecir comportamientos, en grandes volúmenes de información.

Cada una de ellas, posee sus ventajas y desventajas, y la elección de la misma depende en gran medida del objeto de estudio y de la finalidad de la investigación. Para el campo de la seguridad vial, Redes Neuronales Artificiales, Redes Bayesianas y Árboles de Decisión; resultan ser las más difundidas y utilizadas (López G., 2013)

Por este motivo se utilizará para esta investigación Árboles de Decisión (ADD), no obstante dado que la principal limitación de este algoritmo es que las reglas de decisión que se extraen son dependientes de la estructura del árbol, pudiendo existir otros patrones no detectados; se realizará en una etapa previa un análisis exploratorio utilizando la técnica multivariante Analisis de Correspondencia Múltiple y Escalonamiento Multidimensional. Dichas técnicas, permiten ayudar a la comprensión del problema, y a la identificación de las principales variables que intervienen en él.

Se utilizará para el procesamiento de datos el lenguaje de programación “R”, que es un lenguaje especializado en el análisis de datos estadístico y la Minería de Datos.

Se detallan las 24 variables de estudio:

* Día: día de la semana
* Rango horario: día / noche
* Clase de día: laborable/fin de semana
* Franja horaria: mañana (6-13), tarde (14-21) y noche (22-5)
* Comisaria que interviene
* Geolocalización
* Unidades de tránsito: auto/auto, moto/auto, auto/peatón, moto/peatón, etc.
* Condiciones climáticas: despejado/nublado/niebla/llovizna/lluvioso/granizo
* Viento: suave/moderado/fuerte/temporal
* Visibilidad: buena/reducida
* Iluminación: natural/ artificial (buena-regular-mala)/zona de penumbras/nula
* Calzada: asfalto/H° A°/tierra suelta/tierra compacta/ripio/arena/pasto corto/pasto largo
* Estado de la calzada: bueno/regular/malo
* Estado del suelo: limpio/sucio/seco/húmedo/charcos de agua
* Señalización vial si/no horizontal/vertical/ sin incidencia
* Semáforos: si/no funciona si/no Estado: intermitente/continuo
* Edad del conductor
* Tipo de siniestro: leve/grave
* Causalidad: Adelantamiento/ Atropello / Choque / Colisión / Contramano-Falla-Fuga / E-S Garaje / Giro a la derecha/ Giro a la izquierda/ Giro I-U/ Incorpora arteria/ Invasión de carril/ Perdida de control/ Por alcance/ Prioridad/ Roce negativo/ Roce positivo/ Vía semaforizada
* Marca de vehículo
* Modelo de vehículo
* Año de vehículo
* Licencia de conducir: si/no
* Judicializado: si/no

**Resultados**

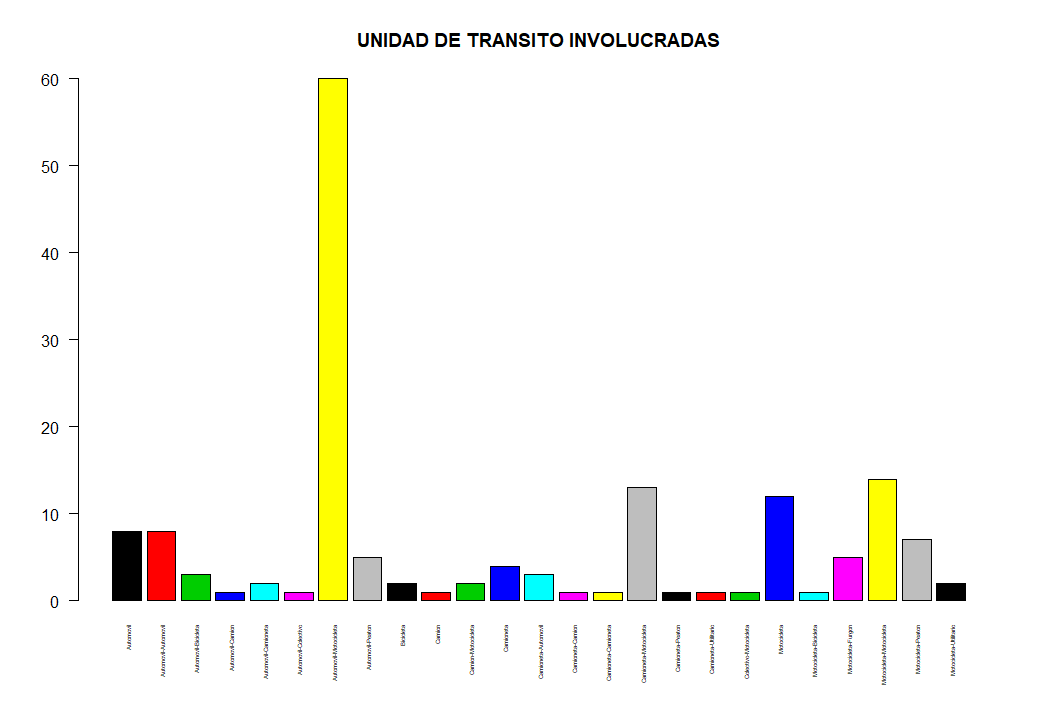
Se analizaron todos los siniestros que transcurrieron en el primer trimestre del 2018 en el departamento de Paraná, donde tuvo intervención la División de Accidentología Vial de la Dirección de Criminalística. A continuación se mencionan aspectos relevantes del análisis descriptivo:

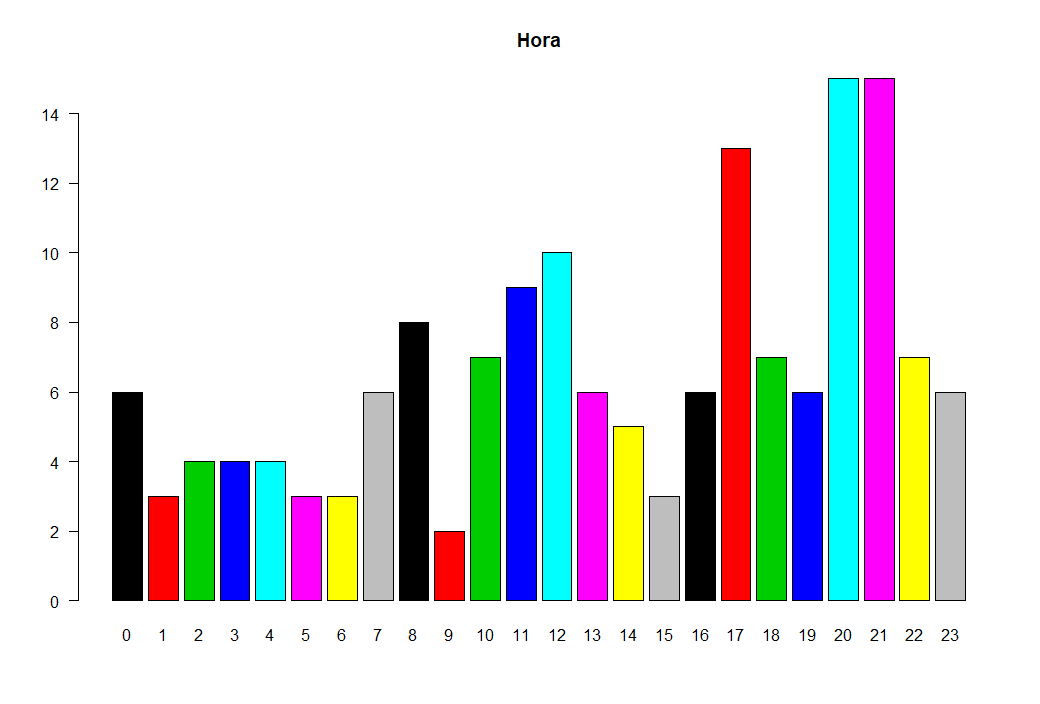
De los 220 casos estudiados, el 33% de los casos revistieron al menos un herido de gravedad, e involucraron a 479 personas. De las personas involucradas, 2% de los accidentados fallecieron, el 14% tuvieron lesiones de gravedad, el 36% lesiones leves, y el 47% sin lesiones. Del total de siniestros el 37% de los casos llegan a judicializarse, correspondiendo el 56% de ellos a siniestros de gravedad.

En el momento del accidente, solamente el 28% de los conductores poseía carnet, y en el 59% de los casos no había señalización vial ni semáforos. Respecto a las edades de los conductores, el 10% son menores de 20 años, el 25% están entre los 21 y 30 años, 39% entre los 31 y los 50 años, y el 26% mayor de 50.

Se puede observar que la mayoría de accidentes que se producen son de automóvil-motocicleta. Registrando la mayor cantidad de siniestros entre las 20 y 22 horas.

Respecto a las condiciones climáticas, el 16% de los accidentes se da en cuando el día es lluvioso o nublado. El 43% transcurre de noche, un 2% es por visibilidad nula, y 20% es porque el estado de la calzada se encuentra en estado malo o regular.





Para el análisis de incidencia de las variables respecto al tipo de lesión se realizó un estudio teniendo en cuenta todos los factores que podían incidir. Se conformaron tablas de contingencia que relacionaran cada una de las variables con la variable dicotómica resultado (siniestro grave/ siniestro leve).

Se contrastó la hipótesis nula que presupone la independencia entre ambas variables, mediante el estadístico Chi-cuadrado para las variables cualitativas politómicas. Para variables dicotómicas se utilizó Chi-cuadrado con la corrección de Yates por ser n≥100; y el Test Exacto de Fisher en el caso de que más del 80% de las frecuencias observadas no fueran mayores que 5, para todos los casos con un nivel alfa de 0,10. Rechazada la hipótesis nula, se midió la intensidad de asociación con el Coeficiente de Contingencia C y V de Cramer, considerando una asociación: pequeño: 0.1, mediano: 0.3, grande: 0.5.

En la tabla 1 se reporta el test estadístico realizado, el valor p, y la conclusión arribada en cuanto al rechazo o aceptación de la hipótesis nula.

Tabla 1. Resumen estadístico de las variables consideradas en el estudio .

(1Chi-cuadrado, 2Test de Fisher), conclusión en cuanto a la hipótesis (\*se rechaza Ho aceptándose que hay evidencia suficiente para concluir que la variable estudiada y el tipo de lesión son estadísticamente dependientes, \*\* No se rechaza Ho) y valores de coeficiente de contingencia y V de Cramer en los casos de rechazo de Ho.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables o factor y test estadístico aplicado** | **Valor p** | **Coeficiente de contingencia** | **V de Cramer** | **Conclusión** |
| 1. Rango horario**1** | 0.07224 | 0.154 | 0.156 | **\*** |
| 2. Franja horaria **2** | 0.005465 | 0.305 | 0.227 | **\*** |
| 3. Unidades involucradas**2** | 0.1099 |  |  | **\*\*** |
| 4 Licencia de conducir **1** | 0.0004052 | 0.297 | 0.312 | **\*** |
| 5. Judicialización del caso**1** | 0.0008801 | 0.268 | 0.279 | **\*** |
| 6. Clase de día**1** | 0.3649 |  |  | **\*\*** |
| 7. Visibilidad**2** | 0.3171 |  |  | **\*\*** |
| 8. Clima**1** | 0.2654 |  |  | **\*\*** |
| 9. Estado de la calzada**1** | 0.2418 |  |  | **\*\*** |

Del análisis se establece que hay dependencia entre la gravedad del accidente, y el rango horario en el que transcurre.  Se encuentra mayor incidencia de siniestros graves de noche que los esperados bajo supuesto de independencia.

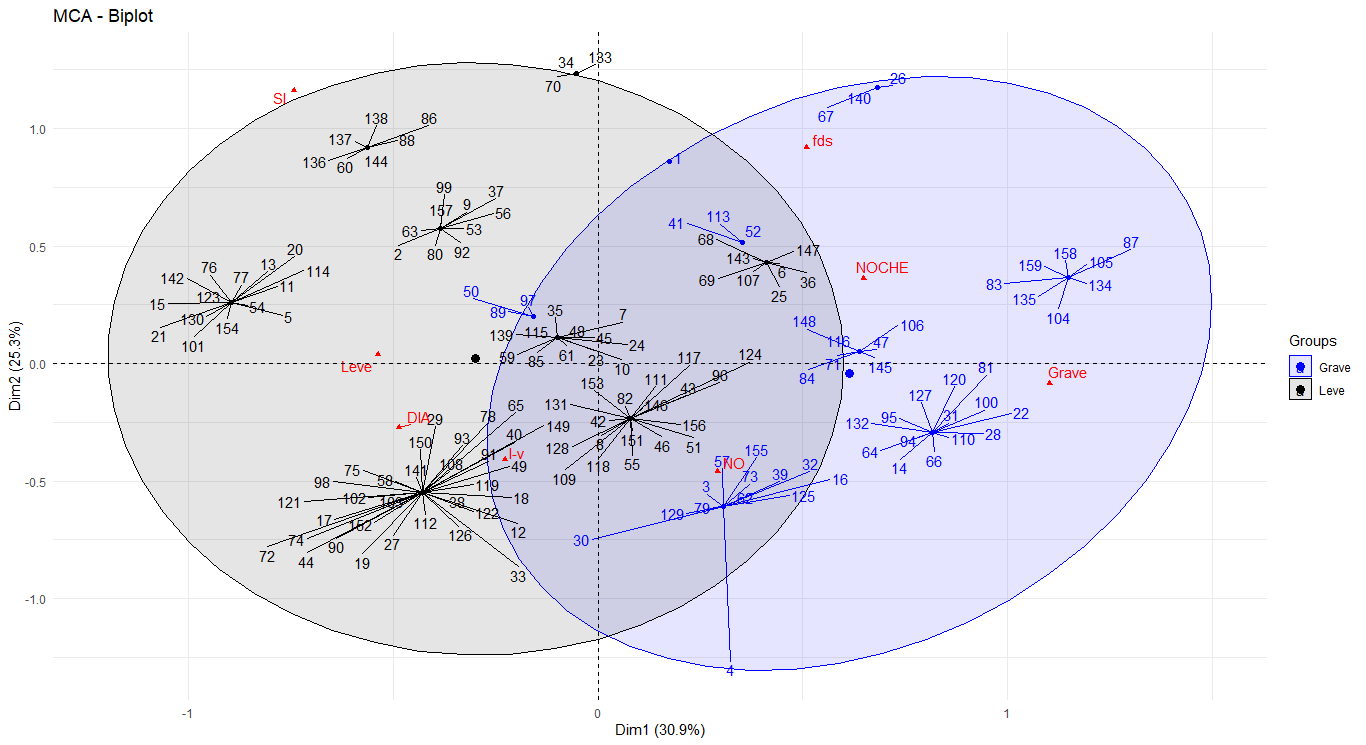
Re categorizando esta variable en tres franjas horarias se observa que hay una asociación entre las siniestros graves entre las 22 y 6 de la mañana, y una asociación entre siniestros leves y el rango horario de la tarde de 14 a 21 horas.

Así también existe una asociación entre tener licencia de conducir y el tipo de siniestro, encontrando una asociación directa entre tener carnet y que el siniestro sea leve, y entre no tener carnet y que el siniestro sea grave.

Se estudió además la asociación entre el tipo de unidades involucradas, encontrando que un accidente de moto con moto, o de auto con auto es menos frecuente de lo esperado, mientras que un accidente de auto con moto es más frecuente que lo esperado bajo el supuesto de independencia.

Resulta llamativo que no se encontró evidencia para rechazar el supuesto de independencia entre la gravedad del siniestro y las unidades involucradas en él.

Para cuantificar el grado de asociación presente entre el conjunto de variables, se utilizó el Análisis de Correspondencia Múltiple. Esta técnica multivariante permitió describir el comportamiento mostrando la interacción entre categorías de las variables, explicando una variabilidad del 56.21% mediante las variables Tipo de siniestro, Rango horario, Clase de día, y Licencia de conducir.



Call:

MCA(X = poison, graph = F)

Eigenvalues

Dim.1 Dim.2 Dim.3 Dim.4

Variance 0.309 0.253 0.242 0.195

% of var. 30.929 25.285 24.238 19.548

Cumulative % of var. 30.929 56.214 80.452 100.000

Categorical variables (eta2)

Dim.1 Dim.2 Dim.3

datos.rango | 0.314 0.099 0.414 |

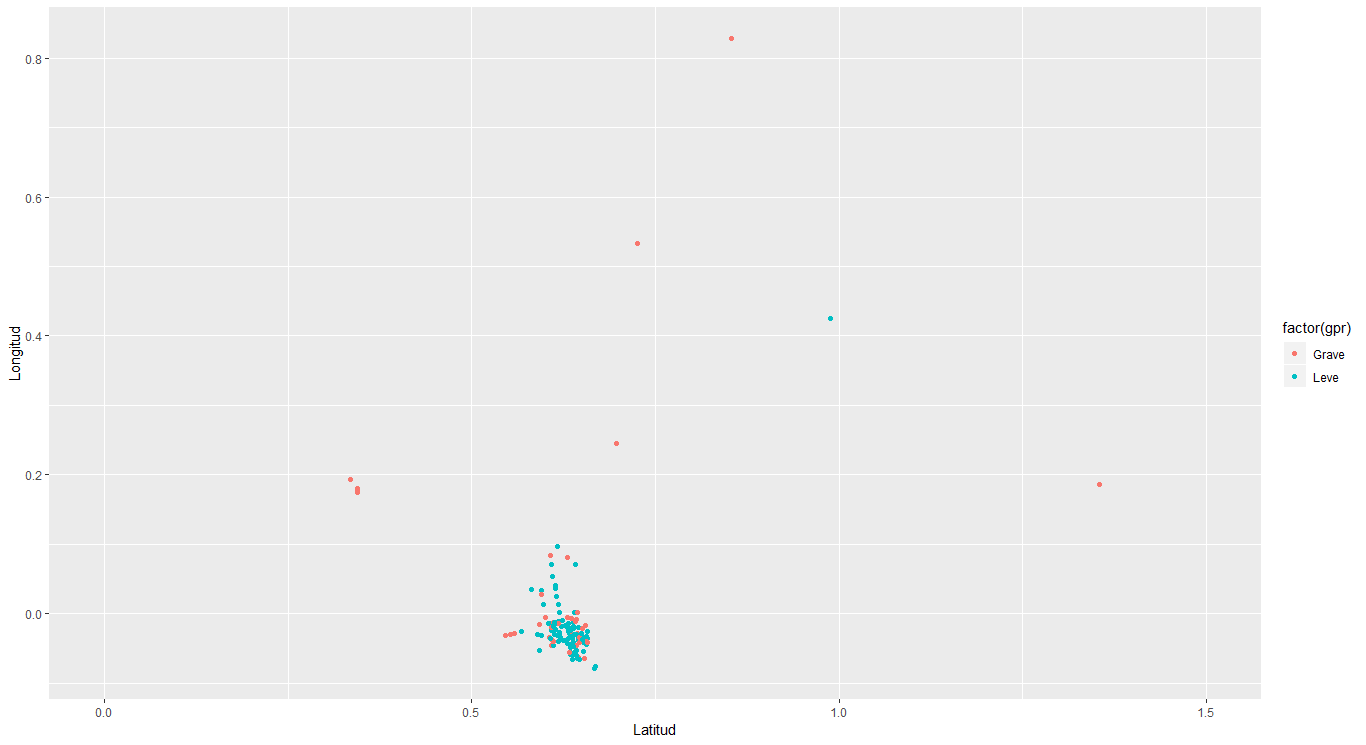
datos.tiposiniestro | 0.591 0.003 0.000 |

datos.carnet | 0.217 0.534 0.089 |

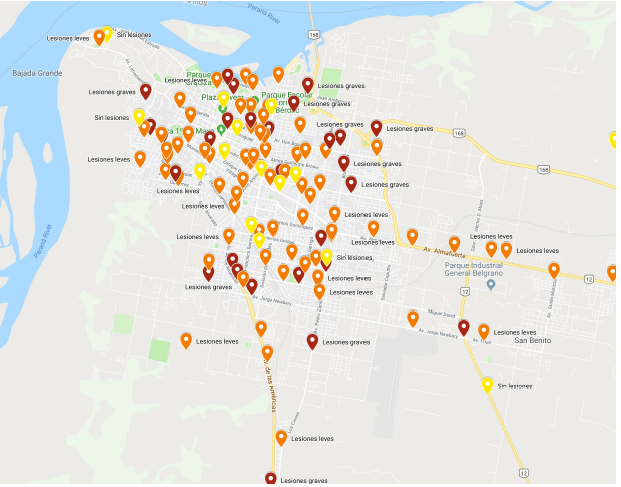
datos.clasedia | 0.116 0.375 0.467 |

Del análisis se observa que existe una asociación entre las categorías siniestro grave, noche, fin de semana, y no tener licencia. Mientras que la categoría siniestros leves queda próxima a la categoría día y tener licencia de conducir.

Para explorar la variable de geolocalización se realizó en primera instancia un Escalonamiento Multidimensional. Esta es una técnica de representación espacial que permite visualizar sobre un mapa un conjunto de objetos cuya posición relativa se desea analizar. En este caso en particular se utilizó las variables latitud, longitud, y el tipo de siniestro. Encontrando que existen ciertas similitudes entre las geolocalizaciones de los siniestros, conformando pequeños cúmulos que resultan destacables dados los pocos casos de estudio que fueron analizados para esta investigación.



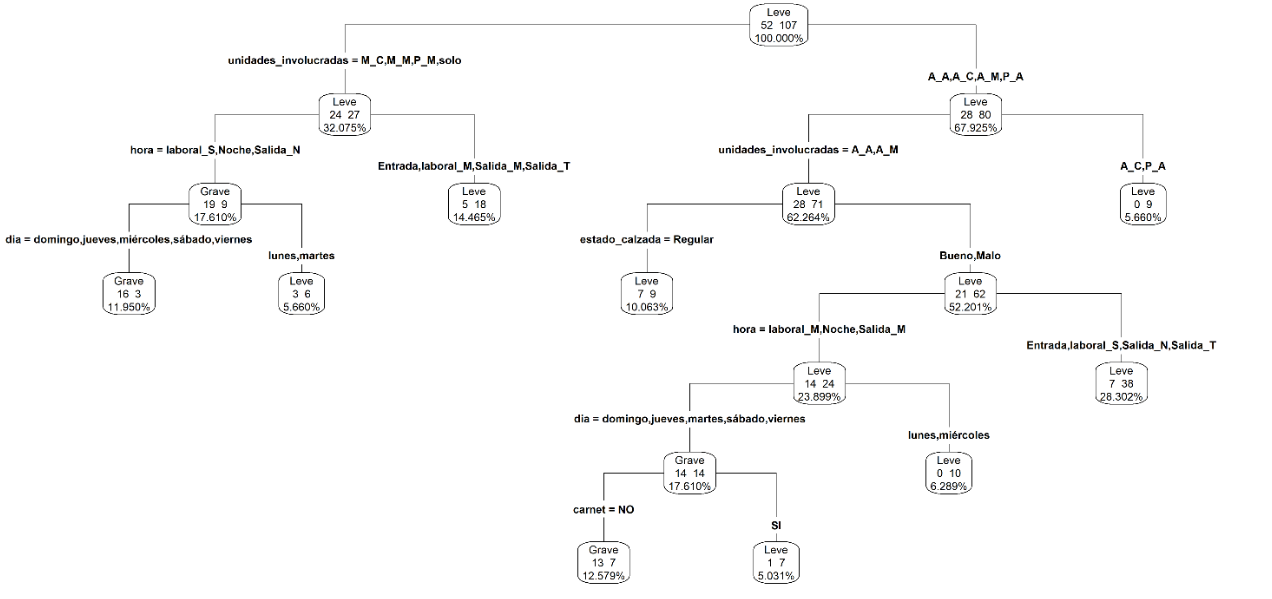
Esta información, luego fue re procesada utilizando el mapa de Paraná, encontrando indicios de que existen calles que resultan más propensas a los siniestros viales como Av. Almafuerte y Av. de las Américas, y también que existen puntos de siniestros superpuestos. No obstante se requieren mayor cantidad de datos para realizar un estudio en profundidad.



Utilizando la misma información se analizaron los siniestros discriminando por el rango horario. Las calles Miguel David, Av. Zanni, Av. de las Américas presentan una concentración de siniestros en horarios nocturnos, mientras que Av. Almirante Brown y Av. Ramirez, Churruarin, presentan concentración de siniestros en horarios diurnos.



Para el análisis conjunto de las 24 variables y su importancia para determinar la gravedad del siniestro se utilizó un Árbol de Decisión (ADD), con la finalidad de encontrar reglas que permiten extraer la información potencialmente útil de las variables y sus categorías.



Observación: Entrada (6-8), laboral\_M (8-12), Salida\_M (13-14), laboral\_S (15-16), Salida\_T (17-19), Salida\_N (20-22), Noche (22-6)

Determinando que:

* Los siniestros que involucran Moto-Moto, Moto-Camión/eta, vehículo solo y se producen entre las 6 y las 14, o entre las 17 y las 19, son siniestros leves.
* Los siniestros que involucran Moto-Moto, Moto-Camión/eta, vehículo solo y se producen entre las 15 y las 16, o entre las 20 y las 6, son siniestros graves.
* Los siniestros que involucran Auto-Moto, Auto-Auto y se producen entre las 9 y las 14 o entre las 22 y las 6, los días lunes o miércoles son siniestros leves.
* Los siniestros que involucran Auto-Moto, Auto-Auto y se producen entre las 9 y las 14 o entre las 22 y las 6, excepto los días lunes o miércoles son siniestros graves.
* Los siniestros que involucran Auto-Peatón, Auto-Camión/eta son siniestros leves.

**Conclusiones**

Como dice la introducción del Anuario Estadístico de siniestralidad vial Año 2017: “En seguridad vial, como en otros ámbitos de la política pública, lo que no se conoce no existe, lo que no se mide no se puede mejorar”.

La severidad de los siniestros viales, y sus consecuencias, nos interpelan a dirigir los esfuerzos a estudiar, analizar y comprender su complejidad, sabiendo que conocer las causales permiten accionar sobre una cantidad de factores que promueven el descenso de estos eventos y la mortalidad que conllevan muchos de ellos.

Este trabajo exploratorio permitió encontrar indicios de asociación entre variables que intervienen en los siniestros, y extraer información que puede ser de utilidad para los profesionales de esta área.

Consideramos que la Minería de Datos es una ciencia altamente valiosa para este tipo de estudios, ya que aporta herramientas para encontrar patrones que no son perceptibles sin ella.

Un trabajo de investigación a gran escala, acompañado de políticas y acciones, podría ser la clave para diseñar estrategias que permita la disminución de siniestros viales y los daños y consecuencias que traen aparejados.

**Bibliografía**

Anuario Estadístico de siniestralidad vial Año 2017. Dirección Nacional de Observatorio Vial.

Bedard, M., Guyatt, G. H., Stones, M. J., y Hireds, J. P. (2002). The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. s.l. Accident analysis and Prevention, págs. 717-727.

de Oña, J., Oqab Mujalli, R., & Calvo, F. (2011). Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accident Analysis & Prevention, 43.

Kashani, A., Mohaymany, A., Ranjbari, A., 2011. A Data Mining Approach to Identify Key Factors of Traffic Injury Severity. Promet-Traffic & Transportation, 23 (1), 11-17.

Kong, C. y Yang, J. (2010). Logistic regression analysis of pedestrian casualty risk in passenger vehicle collisions in China, Accident Analysis and Prevention Volume 42, Issue 4, July.

López Maldonado, G. (2013). Análisis de la Severidad de Accidentes de Tráfico, utilizando Técnicas de Minería de Datos. Granada: Editorial de la Universidad de Granada.

López, G., de Oña, J., & Abellán, J. (2012). Priorización de actuaciones sobre accidentes de tráfico mediante reglas de decisión. Cuadernos Tecnológicos de la Plataforma Tecnológica Española de la Carretera (PTC), 8.

OMS. (2017). OMS | 10 datos sobre la seguridad vial en el mundo. WHO. Consultado el 20 de abril de 2018 en http://www.who.int/features/factfiles/roadsafety/es/#.Wtn4xoyVvLg.

Rivera, J., & Martínez Micakoski, F. (2015). Análisis de variables asociadas en siniestros viales dentro de un área geo referenciada para establecer probabilidades de hospitalización, caso de aplicación.

Rodolfo Nuñez [et.al.]. (2018). Programa de estudios sobre siniestros viales. Ciudad Autónoma de Buenos Aires: Ediciones SAIJ.

Rubio, J., Montt,C. y Lanata ,S. (2013). Análisis de datos para el entrenamiento de redes neuronales para estimación de accidentes de tránsito, actas del Congreso Chilena de Investigación de Operaciones, OPTIMA, paper 59, Concepción, Chile.

Sminkey, L. (2011). Plan mundial para el decenio de acción para la seguridad vial 2011-2020. Ginebra: Naciones Unidas: Decenio de Acción para la Seguridad Vial, 3-25.