

Ingresos a Uruguay

Juan Cadenas, Sol Astengo, Gabriel Bermudez

Introducción

Uruguay se construyó bajo la base de la emigración, y durante su historia el flujo de personas fue determinante a la hora de moldear la estructura de su sociedad. Pero **mas** allá de la población residente, existe un alto porcentaje de individuos que visita el país por lapsos de tiempo **mas** reducidos y motivos diversos, pero que de igual manera inciden directamente en su economía.

Falta referencia temporal de los datos, es un año varios....

Para esto indagaremos sobre los datos ofrecidos por el sistema de datos abiertos, “Turismo Receptivo”. Analizaremos puntualmente a que areas corresponden los gastos de los visitantes, sus distribuciones y de que maneras podemos llegar a explicarlos según el perfil de quien ingresa.

deberían incluir un link a los datos

La base de datos cuenta con **mas** de 59500 observaciones y 48 variables entre las que destacamos, nacionalidad del visitante, motivo de ingreso, duración de la estadía, nivel de educación, ocupación, entre otros.

Datos

Les faltó incluir el tipo de variable y el rango de valores o categorías dependiendo el caso, les quedaron las columnas solapadas en la segunda hoja y no se ve bien.

Tabla 1: Variables de interés y su descripción

nombreDeAtributo	descripcion
Lugar Ingreso	Lugar por donde se ingreso
Pais	Es la Nacionalidad de quien ingresó
Motivo	Es el motivo de ingreso
Ocupacion	Es la ocupación o profesión del turista
Estudio	Es el nivel de estudio del turista
Destino	Es el destino principal
Localidad	
Alojamiento	Es el alojamiento

nombreDeAtributo	descripcion
TransporteLocal	Transporte usado en nuestro pais durante su estadia
Destino	Es la zona de destino principal
Estadia	Dias que permanecio en nuestro pais
Gente	Cantidad de personas del grupo encuestado
GastoTotal	Gasto total, el valor de este gasto es medido en Dólares Americanos Corrientes
GastoAlojamiento	Gasto para el Alojamiento, el valor de este gasto es medido en Dólares Americanos Corrientes
GastoAlimentacion	Gasto para la Alimentación, el valor de este gasto es medido en Dólares Americanos Corrientes
GastoTransporte	Gasto para el Transporte, el valor de este gasto es medido en Dólares Americanos Corrientes
GastoCultural	Gasto para temas Culturales, el valor de este gasto es medido en Dólares Americanos Corrientes
GastoTours	Gasto para Tours, el valor de este gasto es medido en Dólares Americanos Corrientes
GastoCompras	Gasto para Compras, el valor de este gasto es medido en Dólares Americanos Corrientes
GastoOtros	Gasto para otros gastos, el valor de este gasto es medido en Dólares Americanos Corrientes

Objetivo

Buscamos a partir de las variables anteriormente nombradas, entender como influyen en el gasto, si es que lo hacen. Veremos como se comportan los gastos según el motivo de visita al país y si existe alguna relación entre ellos. También tendremos en cuenta a la hora de mirar la forma de estos gastos las nacionalidades correspondientes de los individuos. En **conclusion** nos centraremos en el gasto como nuestra variable de respuesta, y trataremos de explicarlo en función de las características de los no residentes que consumen dentro de nuestra economía.

el gasto es muy general, es gasto total, promedio, diario...en pesos dólares, por eprsona?

Análisis Exploratorio de Datos

Realizamos un análisis exploratorio de los datos, con el fin de familiarizarnos con ellos y entenderlos mejor.

Entre los gráficos y tablas que generamos para visualizar los datos, destacamos los siguientes, que nos ayudan a darnos cuenta cuales serán las mejores formas para cumplir con nuestros objetivos;

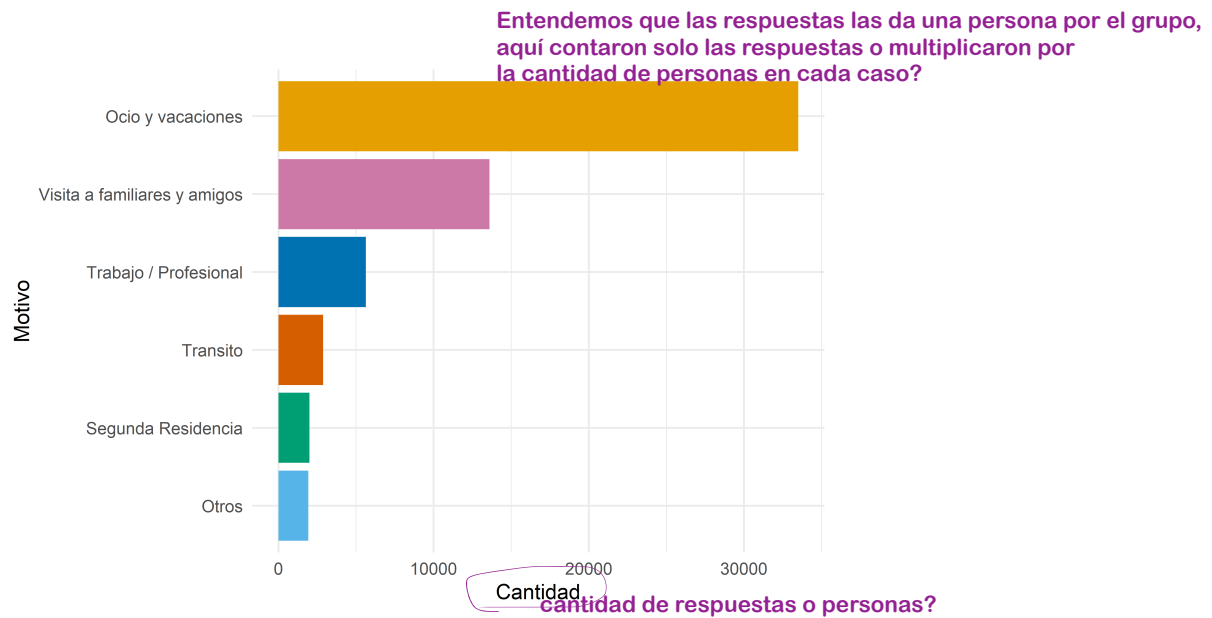


Figura 1: Distribución del motivo de ingreso

Tabla 2: Distribución del MOTIVO de ingreso

Motivo	cant.
Ocio y vacaciones	33498
Visita a familiares y amigos	13597
Trabajo / Profesional	5631
Transito	2876
Segunda Residencia	2008
Otros	925
Estudios	276
Salud / wellness	245
Compras	223
Religioso	176
MICE	75

Importante usar referencias cruzadas en el futuro

Para la figura 1 generalizamos en algunas opciones del motivo del viaje, ya que contaban con pocas observaciones en comparación a otras, las opciones “Estudios”, “Salud”, “Compras”, “Religioso” y “MICE” comienzan a ser parte de la opción “Otros”. En un primer plano podemos apreciar que la gran mayoría de las personas que ingresan al país lo hacen exclusivamente por motivos recreativos y/o de índole social.

Muchas de las observaciones corresponden a un grupo de varios individuos. Trabajamos bajo el supuesto de que todos los integrantes del grupo comparten nacionalidad así como los gastos refieren a la suma de los gastos de todo el grupo.

Siempre recomendamos visualizar antes de tablas!
Mismo en la tabla podrían haber agrupado las categorías
menos importantes. No es relevante cada lugar
sinó los más frecuentes en su cuento.

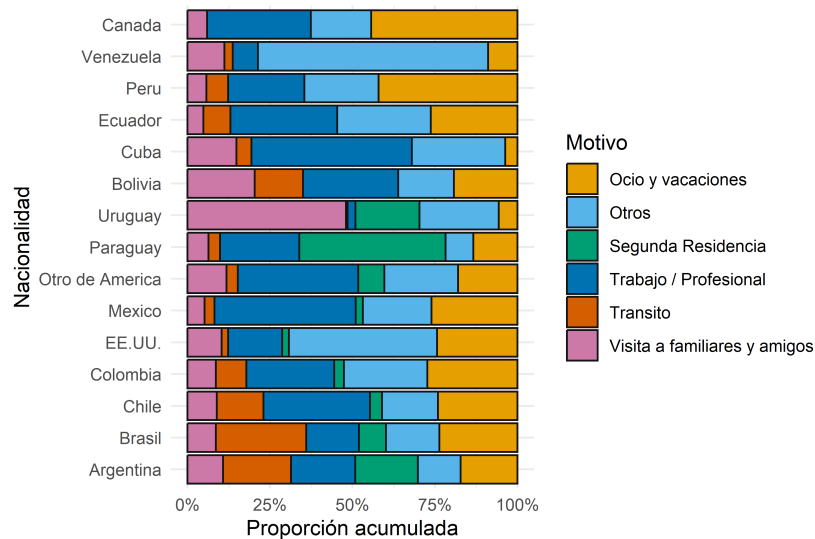
Distribucion de las nacionalidades de los visitantes

Tabla 3: Cantidad de ingresos por nacionalidad del ingresante

IdNacionalidad	Pais	cant.
10	Argentina	29832
19	Brasil	11168
1	Uruguay	9522
40	Chile	1561
30	Paraguay	1310
49	Otro de America	818
70	EE.UU.	804
51	España	648
59	Otro de Europa	642
35	Peru	424
32	Colombia	412
36	Venezuela	341
42	Mexico	276
55	Alemania	263
52	Italia	233
53	Francia	216
33	Ecuador	198
41	Canada	150
54	Gran Bretaña	120
18	Bolivia	107
69	Otro pais de Asia	100
79	Africa u Oceania	91
57	Suiza	76
61	Israel	65
43	Cuba	49
56	Suecia	49
62	China	28
63	Japon	27

Suponemos que gran parte de los uruguayos que vuelven a pasar unos días es con propósitos sociales, pero y el alto porcentaje de Argentinos? Brasileños? a que vienen? El siguiente gráfico nos muestra los motivos de visita para las personas con nacionalidades americanas.

Es importante escribir correctamente, tienen varias faltas además de este tipo de frases que no están bien.



deberían acalrar que es solo los de America del norte y sur en el título

Figura 2: Distribución de motivos de visita según nacionalidad

Agregar el tipo de gráfico en el título, grafico de barras apiladas al 100% de motivo de la visita según nacionalidad

Respondiendo a las preguntas anteriores, personas de Argentina y Brasil ingresan al país por motivos varios, sin uno que predomine mucho sobre los demás motivos. En cambio, gracias a este gráfico podemos visualizar comportamientos extraños cuando se trata de otros países, por ejemplo el motivo que predomina para que los venezolanos ingresen al país es “Otros”, tomando en cuenta la situación política y económica del país tendría sentido que se clasifique en esta categoría sus casos.

En la figura 2 vemos únicamente la nacionalidad de los ingresantes americanos. Que tal se comportaran los datos si se trata de nacionalidades europeas o asiáticas?, para responder esa pregunta realizamos otro gráfico igual, esta vez incluyendo los países deseados.

Deben escribir con mayor formalismo en un trabajo no es un mensaje de Whatapp.

Algo así:

“Nos interesa analizar la distribución del motivo de la visita según nacionalidades europeas y asiáticas.

¿Hay diferencias en los

motivos para estas nacionalidades respecto a las analizadas en la Figura 2?”

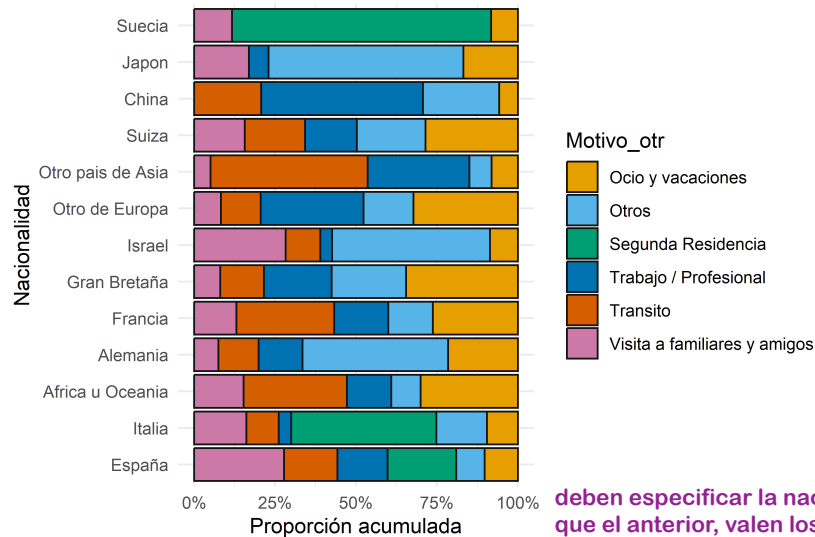


Figura 3: Distribución de motivos de visita según nacionalidad

Observamos que en general las personas ingresantes al país con nacionalidades europeas o asiáticas, lo hacen con diferentes fines, sin uno que predomine sobre otros. Excepto cuando se trata de las personas provenientes de Suecia: este caso llama la atención, pero la predominancia de integrantes con segunda residencia, en realidad se da por la generalización que hicimos antes en la que agrupamos los motivos menos frecuentes y porque contamos con muy pocas observaciones provenientes de suecia en nuestra base de datos.

Siendo esta una variable relevante en su trabajo no queda claro cómo la consideraron.

Si es en el total de la estadía no tiene sentido para la comparación, ni tampoco el total de gasto para el grupo.

Importante si es el precios corrientes, constantes. Dólares o pesos...

Gasto en alojamiento

Para prever como podría comportarse el gasto en alojamiento, podemos ver cuáles son las opciones de este más usuales entre los ingresantes. Para esto se genera la siguiente tabla:

Tabla 4: Cantidad de observaciones por alojamiento

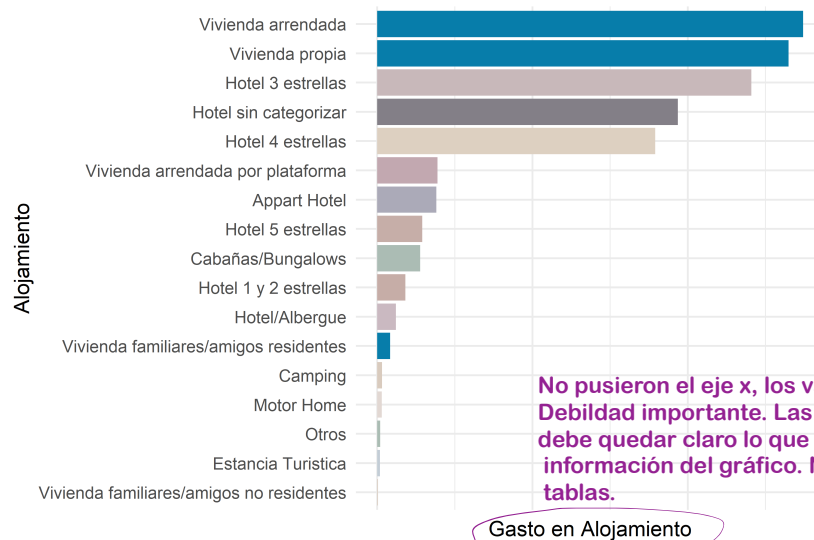
TABLAS....

Alojamiento	cant.
Vivienda familiares/amigos residentes	17946
Hotel 3 estrellas	8623
Hotel sin categorizar	7136
Vivienda arrendada	4476
Hotel 4 estrellas	4417
Ninguno	4187
Vivienda propia	3553
Vivienda familiares/amigos no residentes	1497

Alojamiento	cant.
Hotel 1 y 2 estrellas	1430
Hotel/Albergue	1332
Cabañas/Bungalows	1321
Vivienda arrendada por plataforma	822
Appart Hotel	729
Hotel 5 estrellas	712
Camping	511
Otros	434
Motor Home	180
Tiempo Compartido	92
Barco, Yate, Crucero	47
Estancia Turística	42
Bed y Breakfast	34
Sin Datos	9

mejorar redacción....

En la tabla 4 podemos ver que la forma de alojamiento que es **mas** usual entre los visitantes es la vivienda de familiares/amigos que son residentes en el país, por lo que podemos asumir que estas personas **gastaran** poco o nada en alojamiento?, veamoslo



No pusieron el eje x, los valores del gasto en que unidades. Debilidad importante. Las figuras deben ser autocontenidas, debe quedar claro lo que están dibujando cuando leemos la información del gráfico. No pasa en la mayoría de sus gráficos y tablas.

Gasto en Alojamiento

Figura 4: Gasto en alojamiento según alojamiento

En la figura 4 vemos como las personas que pertenecen al grupo de “vivienda familiares/amigos residentes” cuando se trata del alojamiento son de los que menos gastan en este, por lo tanto podemos decir que el supuesto anterior se cumple. En este gráfico también llama la atención

que las personas que ingresan al país y cuentan con vivienda propia son de las personas que mas gastan en alojamiento, por lo que podríamos suponer que estas personas se encuentran en proceso de pagar tal vivienda.

Gasto en Transporte

De igual forma que el gasto en alojamiento esta directamente relacionado con que forma de alojarse en el país, en su estadía, elige cada visitante, el gasto en transporte se puede relacionar con cual es el método de transporte local **mas** usado por los ingresantes. Para visualizar como podría comportarse el gasto en transporte, realizamos una tabla en donde vemos **cual** es el **mas us**ualmente usado entre los ingresantes.

Tabla 5: Cantidad de observaciones por transporte utilizado localmente

TransporteLocal	cant.
Auto propio	30648
Taxi - Bus	14640
Auto familiares / amigos	7580
Auto alquilado	2524
Ninguno	2306
Otros	1801
Sin Datos	31

De esta forma en la tabla 5 podemos ver que gran parte de los ingresantes utiliza auto propio para movilizarse dentro del país. **Pero esto significa que el gasto en transporte es 0, o se toma en cuenta el gasto en gasolina, si el auto necesitaría alguna reparación en el viaje, entre otros factores?.** Para responder esta pregunta realizamos el siguiente gráfico

La pregunta no está bien formulada, sigo marcando el tema de la escritura porque es muy importante mejorar este tema en trabajos futuros.

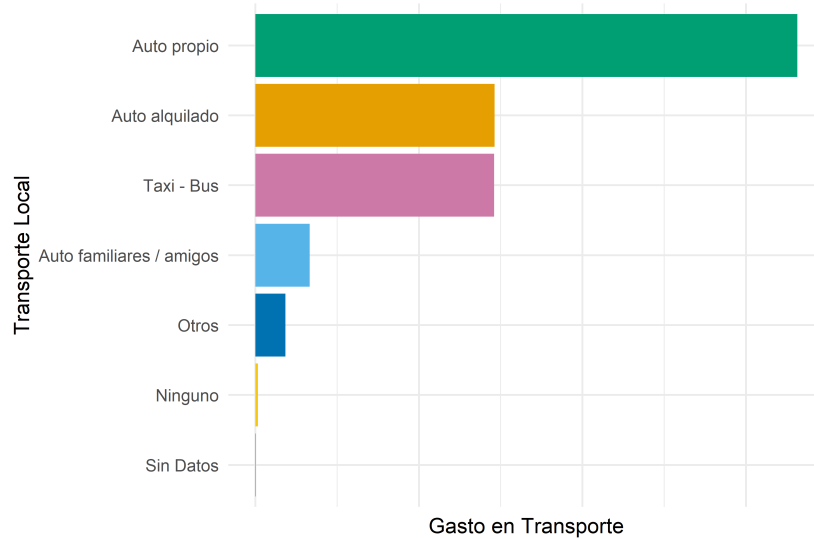


Figura 5: Gasto en transporte según transporte local

En la Figura 5 podemos ver que todos los gastos relacionados a auto propio como transporte local, parecen estar incluidos en la variable de gasto en transporte.

recién aparece algo de la periodicidad de el período analizado, deberían haberlo incorporado antes en la descripción de datos que quedó pobre

Nota sobre fechas

Se identificó un posible error en la fecha de ingreso del 2011-12-27, con una fecha de egreso 8 años después. Se supone que en vez de 2011, debería ser 2018, ya que la duración de estadía es de 14 días. Por lo tanto, la fecha de ingreso se ha corregido de 2011-12-27 a 2018-12-27 para reflejar correctamente la duración de la estadía.

Incorporación de variables gastos por persona.

Para poder conseguir los objetivos deseados decidimos que lo mejor es tener los gastos divididos por persona, ya que tenemos una variable que describe la cantidad de personas que se encontraban en el grupo de viaje y suponemos que los gastos son los realizados entre todos los del grupo en conjunto. Así que agregamos a nuestros datos las columnas correspondientes a los gastos por persona para facilitar cálculos posteriores. (almacenamos los cambios en “receptivo_fecha”). A continuación en la tabla 6 podemos ver algunas de las variables creadas

Tabla 6: Algunas nuevas columnas de gastos por persona

GastoCultural_porPe	GastoTours	GastoTours_porPe	GastoCompr	GastoCompras_porPe	GastoOtros
60.00	0	0	785.00	157.00	60.00
57.06	30	30	48.69	48.69	2.48
137.00	0	0	500.00	125.00	1320.00
0.00	0	0	210.00	70.00	0.00
0.00	0	0	0.00	0.00	0.00
0.00	0	0	0.00	0.00	30.00
0.00	0	0	176.84	44.21	15.05
0.00	0	0	140.00	70.00	66.00
0.00	0	0	0.00	0.00	0.00
80.00	0	0	0.00	0.00	110.00
80.00	0	0	140.00	70.00	154.00

¿Como cumplir con el objetivo?

Para cumplir con lo deseado, utilizaremos una serie de herramientas, comenzaremos creando una Shiny app interactiva que permitirá explorar y manipular estos datos. A continuación, utilizaremos herramientas de mapeo para visualizar geográficamente la información sobre los gastos realizados por los ingresantes en Uruguay. Luego, implementaremos métodos gráficos para visualizar datos de dos o más variables cualitativas, facilitando la interpretación y el análisis de relaciones complejas entre variables como la nacionalidad y el motivo de ingreso, con respecto a los gastos. Finalmente, implementaremos modelos de árboles y bosques para realizar predicciones precisas basadas en patrones identificados en los datos.

Shiny App

Para permitir una visualización interactiva de cómo los visitantes gastan en diversas áreas, optamos por una serie de gráficos en los cuales también se tomaran en cuenta otras variables, como el motivo del viaje, la nacionalidad del visitante, el nivel de educación de este, entre otras, y de esta forma podremos ver como influyen las demás variables en los gastos en diferentes areas.

Nuestra app consta con varias pestañas, cada una asignada a una variable de interés, dentro de la cual se despliega un gráfico de distribución del gasto según el valor de la variable que deseemos ver. Al ser la base de datos bastante pesada y quizá también el código de la shiny, al querer subir la app a shinyapps.io nos devuelve un error de almacenamiento haciendo referencia a que consume mucha memoria. Intentamos reducir la base de datos considerablemente y hasta quitar pestañas para hacer más liviano el código, pero sin resultado. Creemos que seguirle

quitando información y código buscando la posibilidad de subirla haría que en el proceso se perdiera la capacidad de esta de reflejar correctamente la distribución que nos interesa. (La app está disponible en el repositorio de GitHub, carpeta “app proyecto”)

Mapa de Uruguay

Nos interesa observar como distribuye el gasto por persona segun nacionalidades (solo americanas) de los turistas.

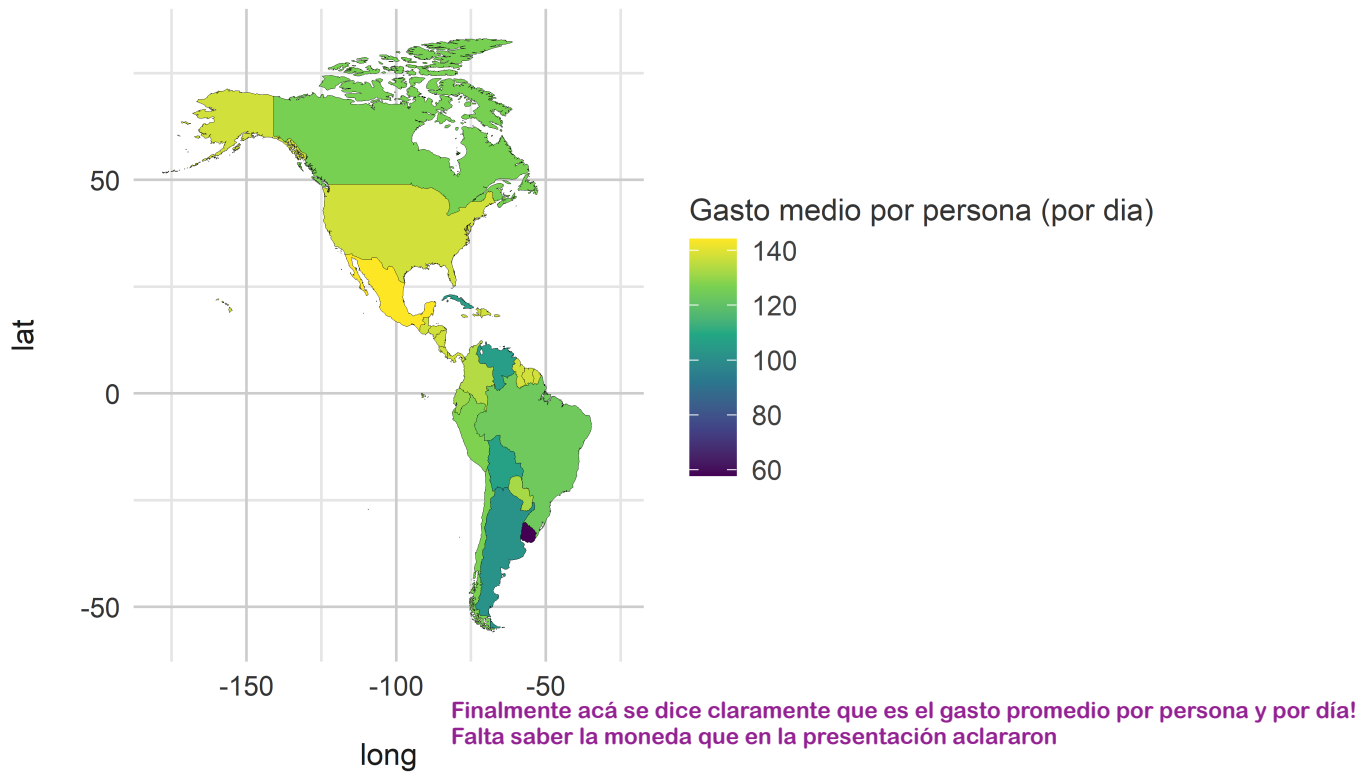


Figura 6: Mapa de América Latina: Gasto Promedio por Persona por Día.

Nos gustaría resaltar que para la creación de este mapa debimos generalizar un poco para una mejor visualización de los datos en el mapa, ya que no contamos con datos de algunos países.

En la figura 6 podemos ver como se distribuye el gasto por persona por **día segun** la nacionalidad de la persona que ingresa al país, tomando en cuenta solo las nacionalidades americanas, ya que sabemos que la mayor parte de los ingresantes son de tales nacionalidades. En el gráfico se logra visualizar que al parecer las personas con nacionalidad mexicana son las que más gastan en el país cuando ingresan, y que el gasto promedio de las personas de nacionalidad

uruguay al ingresar al país es el más bajo, esto ultimo tiene sentido, ya que podemos ver en la shiny app que las personas uruguayas que ingresan gastan más que nada en alimentacion, por lo que podemos suponer que gastos como el alojamiento lo tienen mayormente cubiertos por familiares o amigos que visiten. En cambio es un poco raro que las personas que mas gasten sean exactamente las mexicanas, en la shiny app podemos ver que las personas de este país gastan mayormente en alojamiento y alimentacion, pero esto no es algo extraño ya que la mayoría de los ingresantes sin importar su nacionalidad (exceptuando Uruguay), gastan mas que nada en esas areas.

Otros gráficos

Para intentar explicar el porque las personas con nacionalidades mexicanas gastan **mas** cuando ingresan al pais podemos ver un **grafico** que compara el gasto medio por persona por **dia**, en caso de nacionalidad Mexicana y Uruguay **segun** el motivo, dejando solo los motivos con mayores observaciones para una mejor visualizacion de los datos.

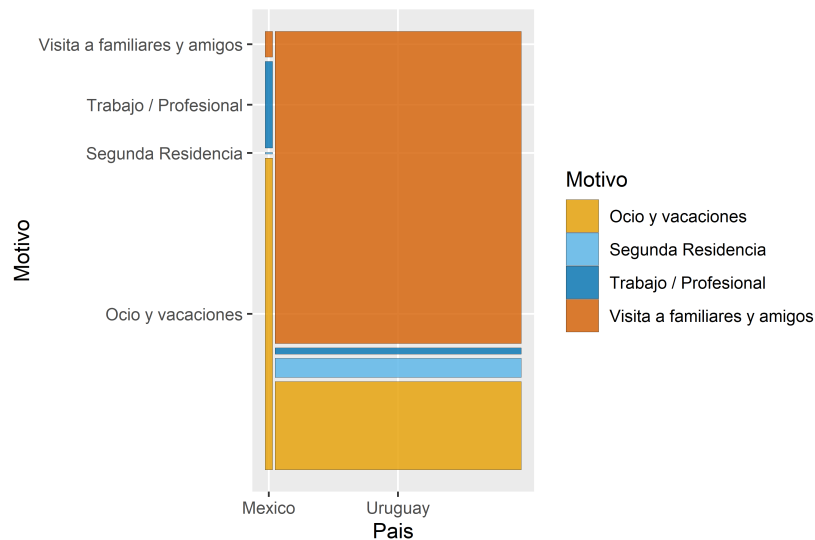


Figura 7: Comparacion Mexico-Uruguay, en gasto **segun** motivo

Ahora podemos hacer otro **grafico** para ver en que se diferencian los gastos de los turistas mexicanos a los de los uruguayos.

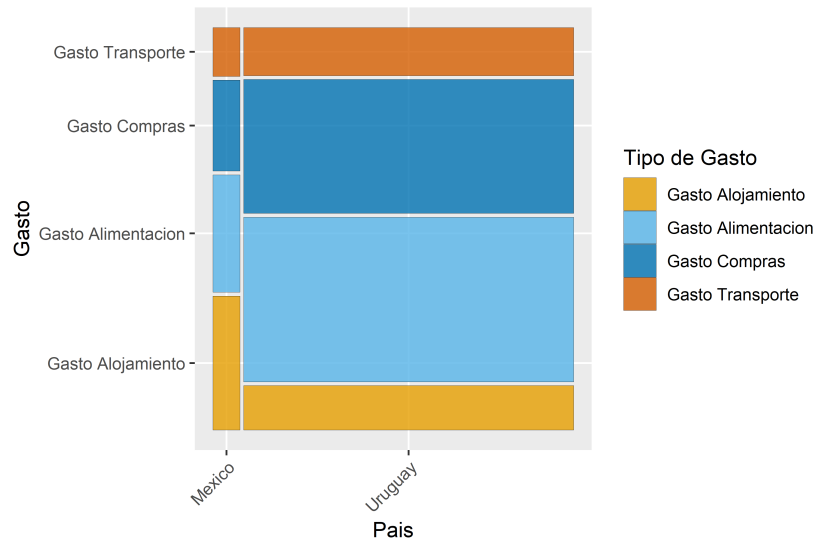


Figura 8: Comparación de los tipos de gastos por persona por día entre turistas mexicanos y uruguayos.

Viendo en conjunto la figura 7 y 8, nos damos cuenta de una clara diferencia en el motivo del viaje, los mexicanos ingresan al país mayormente con el fin de vacacionar, en cambio los uruguayos lo hacen para visitar a sus familiares y amigos. También vemos que en el caso de los mexicanos gastan más del doble en alojamiento de lo que gastan los uruguayos. Por lo tanto podríamos concluir que el motivo por el que los ingresantes con nacionalidad mexicana sean los que mas gastan en el país de toda america, es que entran al país por ocio y vacaciones y gastan mucho en alojamiento.

Arboles y Bosques

Intentaremos generar un árbol de decision y un bosque aleatorio, con el fin de predecir el gasto total por persona segun las demas variables con las que contamos.

En el análisis realizado, utilizamos un árbol de decisión y un bosque aleatorio para predecir el gasto total por persona (GastoTotal_porPersona) de los visitantes en Uruguay. Sin embargo, nos dimos cuenta de que ambos modelos no logran captar bien los valores reales, y suelen subestimarlos. Uno de los problemas principales es que el error cuadrático medio (RMSE) está en unidades al cuadrado de la variable original, lo que hace que el error parezca más grande y difícil de interpretar.

no es correcto, la interpretacion del RMSE queda en las unidades de la variable de interés

Durante el curso, no se nos presentaron las herramientas necesarias para solucionar estos errores y mejorar el modelo. Así que, básicamente, el modelo “da mal y ya”. Para solucionar esto en el futuro, sería útil tener más variables o diferentes tipos de variables, pero eso escapa del alcance

Aunque el modelo les de mal tendrían que haber resumido los resultados de lo que hicieron un modelo malo también es un resultado y hay veces que se requiere mucho esfuerzo para mejorarlo no solamente conocimiento. No es una exigencia del curso que el modelo les de bien ya que vimos muy poco pero si poder resumir lo que exploraron e hicieron.

del curso y los datos que tenemos ahora. Por estas razones, los modelos desarrollados no son adecuados para una predicción precisa en este contexto.

Paquetes utilizados y fuente de los datos

Datos:

Catalogo de Datos Abiertos. Ministerio de Turismo. <https://catalogodatos.gub.uy/dataset/ministerio-de-turismo-turismo-receptivo>

Paquetes:

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>. Jeppson H, Hofmann H, Cook D (2021). *ggmosaic: Mosaic Plots in the ‘ggplot2’ Framework*. R package version 0.3.3, <https://CRAN.R-project.org/package=ggmosaic>. Therneau T, Atkinson B (2023). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1.23, <https://CRAN.R-project.org/package=rpart>. A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18–22.

Como comentario final, recorrieron casi todas las partes del proyecto aunque les faltó describir los resultados. Se que el modelo les dió mal pero es un resultado y el curso no tenía el foco en obtener el mejor modelo sino saber plantear un modelo definiendo que variable es de interés predecir y cuales son posibles predictoras.

Me gustó su presentación en clase y el trabajo que hicieron con la aplicación, lograron hacer andar la reactividad!

Como les indiqué a lo largo del trabajo tienen que mejorar la escritura además el tema de la descripción de los datos y las visualizaciones que hicieron no están autocontenidas, no queda clara la información que muestran en la misma.

Espero sigan aprendiendo a lo largo de otros cursos de la Lic. en estadística y fortaleciendo algunas habilidades más.

Puntos totales 85/100