

The background of the slide is a photograph of a wide, calm lake. In the foreground, the water is a deep blue with gentle ripples. A line of tall, golden-brown reeds or grasses stretches across the middle ground. In the distance, a thin line of green trees is visible against a pale, clear sky. A small, bright white spot, possibly a moon or a reflection, is visible in the upper right portion of the sky.

**IV Jornadas de Estadística Aplicada
La Paloma**

**Turistas domésticos en el Departamento de
Rocha. Clasificación mediante técnicas de
aprendizaje estadístico**

Gonzalo Perera - Martin Sprechmann - Mathias Bourel

Temáticas a abordar

- **Motivación y objetivos**
- **El Departamento de Rocha**
- **La segmentación de partida**
- **Resultados**
- **Conclusiones**

Motivación y objetivos

Motivaciones

- **La importancia del turismo doméstico para el Departamento de Rocha.**
- **Complementar la segmentación con una clasificación de los turistas.**

Objetivo:

Clasificar un nuevo individuo que no ha contestado el cuestionario de partida en uno de los cuatro clusters obtenidos en una segmentación.

El Departamento de Rocha

- **Aproximadamente 70.000 habitantes.**
- **170 kilómetros de Costa Atlántica.**
- **Ambiente relativamente prístino.**
- **Turistas domésticos representan 67% de arribos (OTR, Enero 2016).**

Recursos y atractivos:

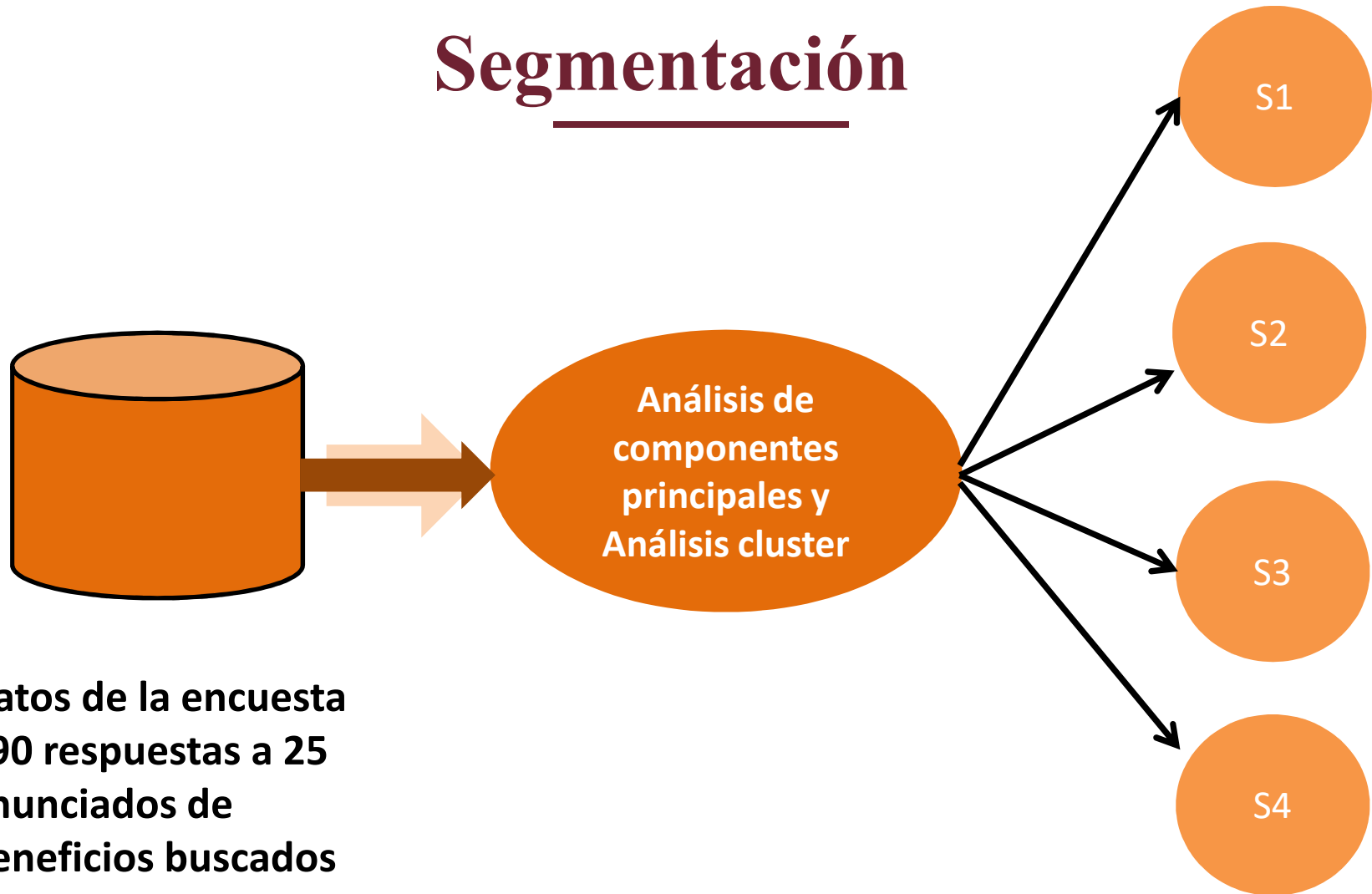
- **Playas, pequeños pueblos costeros, dunas, áreas naturales, palmares y sitios históricos**

La segmentación de partida

- **Se realizó una encuesta online a personas residentes en Montevideo.**
- **Que hubieran pasado al menos un día en el verano de 2013 en el Departamento de Rocha.**
- **25 enunciados de beneficios buscados.**
- **Escala de Likert de 5 puntos.**
- **Se obtuvieron 290 respuestas válidas.**

- **Análisis factorial.**
- **Análisis cluster jerárquico y análisis cluster k-medias**

Segmentación



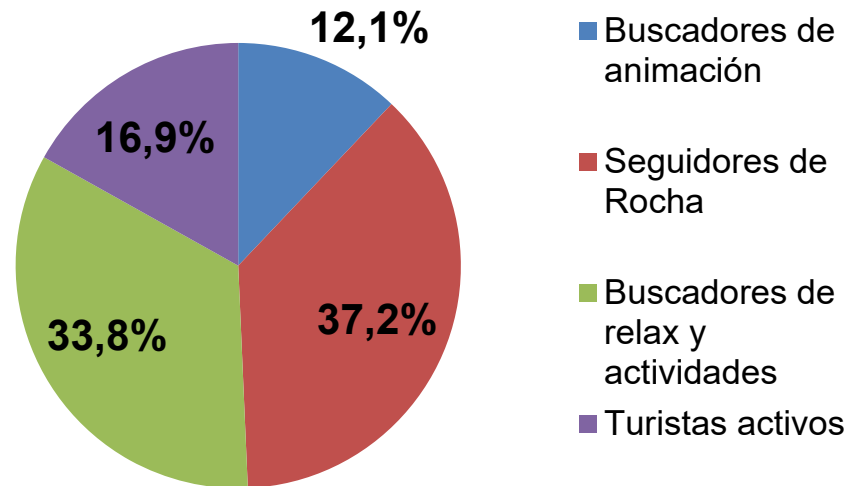
La segmentación de partida

	Buscadores de animación	Seguidores de Rocha	Buscadores de relax y actividades	Turistas activos
Porcentaje sobre el total de la muestra	12,1%	37,2%	33,8%	16,9%

Resultados - Segmentos

Factores	Buscadores de animación n=35 12.1%	Seguidores de Rocha n=108 37.2%	Buscadores de relax y actividades n=98 33.8%	Turistas activos n=49 16.9%
Animación	.430	-.029	-0.008	-.227
Características de Rocha	-.313	.830	-.737	-.132
Relax	-1.990	.263	.328	.184
Naturaleza	-.270	-.045	-.111	.514
Deportes	-.140	-.166	-.222	.909
Actividades y paseos	-.180	-.357	.550	-.184
Camping	.125	-.347	-.365	1.405

Resultados – segmentos obtenidos



- **Buscadores de animación. 12,1%.** Parte de la “**movida joven**”, no toda.
- **Seguidores de Rocha. 37,2%.** **Características de Rocha y relax.** Presumiblemente turismo más tradicional.
- **Buscadores de relax y actividades. 33,8%.** **Actividades paseos y relax.** Capta movimiento de compras en el Departamento.
- **Turistas activos. 16,9%.** **Actividades al aire libre.** Surf, pesca, deportes, lugares ecoturísticos, camping y relax.

Segmento 1 - Buscadores de animación



Buscadores de animación

- Relativamente **pequeño**. **12.1 %** de la muestra
- Mayormente **hombres**. **57,1%**
- **Punta del Diablo** (34,3%) y en menor medida La Paloma y La Pedrera



- Segmento **más joven** en términos relativos.
- Viajan fundamentalmente en **grupos de amigos** (48,6%).
- Se toman **menos tiempo** para **planificar el viaje**.
- Tienen por mucho las **estadias más cortas**.

Segmento 2 – Seguidores de Rocha



Seguidores de Rocha

- El más **grande**. **37.2 %** de la muestra.
- Mayoría de **mujeres** (**59,3%**).



- **La Paloma Grande** (35.2%) y en menor medida La Pedrera y alrededores (15,7%) y Cabo Polonio (13,0%).
- **Marcada importancia de los nodos de La Paloma, La Pedrera y el Polonio.**
- Son los que más viajan en **familia** (50,9%) y los que más alquilan casa en familia (38,0%).
- Presentan las **estadias relativamente más largas.**

Buscadores de relax y actividades

- También **grande**. **33.8 %** de la muestra.
- Equilibrado entre **hombres** y **mujeres**



- Un tercio va al **Este** de **Rocha**. **La Paloma Grande** (26,5%) **Polonio y vecindades** (19,4%), **Punta del Diablo** (15,4%) y **Barra de Chuy** (13,3%).
- Factor **proximidad** a centros de compras?
- Se quedan **más** en **casa propia** que el resto.
- Presentan estadías **relativamente cortas**.
- También van mucho en **grupos familiares** (43,9%).

Segmento 4 – Turistas activos



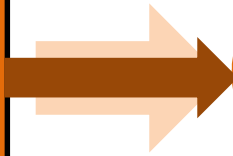
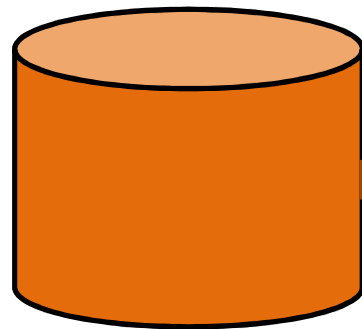
Turistas activos

- Es un cluster **pequeño**. **16.9 %** de la muestra.
- Más miembros son **mujeres** (**55,1%**)



- El **69,4%** de los Turistas Activos tiene **menos de 39 años**.
- Lugares preferidos: **Santa Teresa** (20,5%) y **Punta Rubia** (18,4%).
- Se alojan en gran medida en **Campings** (37,3%)
- Hacen más uso del **ómnibus de línea** que el resto de los clusters (34,8%).

Clasificación



Análisis de
componentes
principales y
Análisis cluster



Desarrollo y
evaluación de
la performance
de los
clasificadores

Datos de la encuesta
290 respuestas a 25
enunciados de
beneficios buscados

Clasificación

- **Se entrenaron modelos de clasificación mediante tres técnicas: Regresión Logística, Máquinas de Soporte Vectorial y Random Forest.**
- **El input del modelo fueron las 290 respuestas de los encuestados a los 25 enunciados.**
- **La etiqueta del modelo es el cluster al que pertenece cada encuestado.**

Clasificación

- **El conjunto de entrenamiento corresponde a dos tercios de la muestra.**
- **El tercio restante se utilizó para verificar la capacidad predictiva del modelo.**
- **Se realizaron varias réplicas de división aleatoria.**

Resultados de la clasificación

	Regresión Multinomial	Random Forests	SVM
Error	48,69%	21,46%	15,65%

Conclusiones

- **Support vector machines es el clasificador con el mejor desempeño para este problema.**
- **Le siguen Random Forest y la Regresión Logística.**
- **El modelo de support vector machines permite validar la solución de clustering k-medias.**
- **El modelo permite clasificar turistas que no fueron muestreados (argentinos, brasileños, etc.).**

Conclusiones

- **La gran mayoría de los estudios de segmentación no utilizan herramientas de machine learning para clasificar a los individuos.**
- **(Dutta et al., 2017) Segmentación y clasificación de turistas domésticos en la India: una perspectiva de stakeholders.**

Trabajo a futuro

- **Utilizar métodos que asignen importancia a las variables explicativas, como CART.**
- **O más aun, métodos de consenso, que combinan métodos de diversa naturaleza.**
- **Muestra independiente de la utilizada en este trabajo.**