



# Exploratory Analysis

*Youpele Michael*

# OVERVIEW

DATASET

EXPLORATORY ANALYSIS

HYPOTHESIS

DETECTED SEASONALITY

LEAD TIME VS CANCELLATION

# DATASET

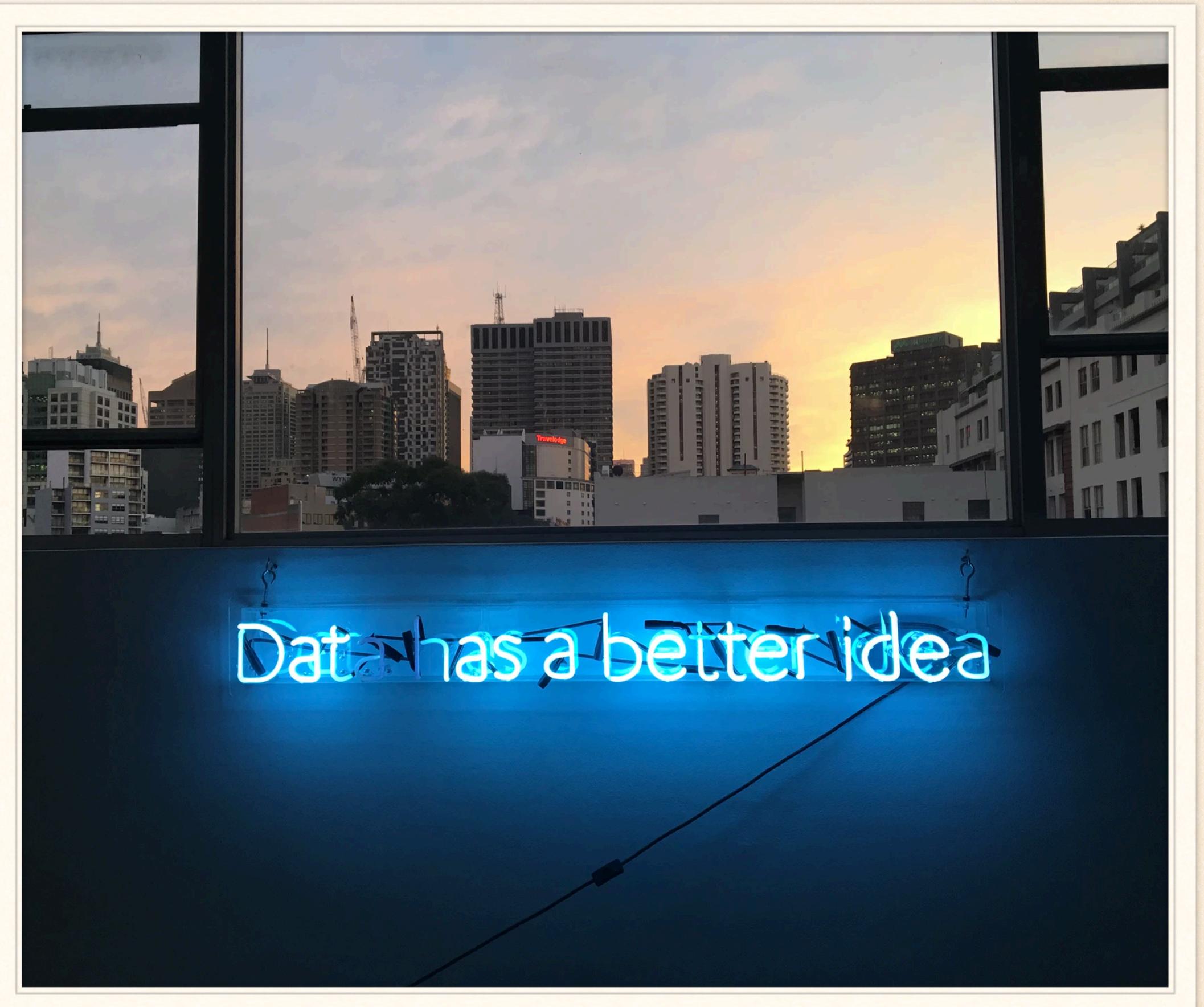
---

INFO  
CLEANING



# DATASET INFO

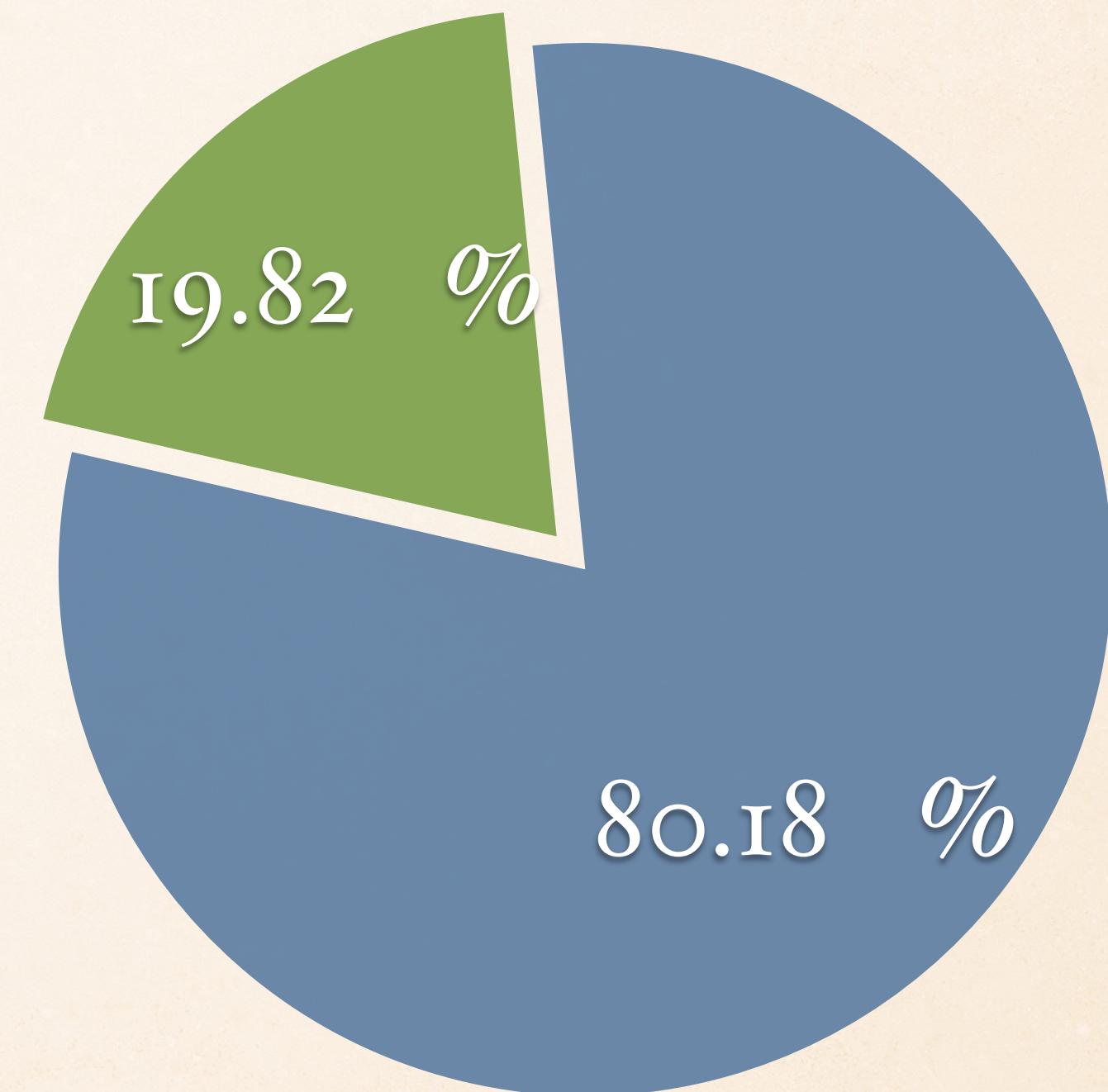
- ❖ 119,390 rows and 32 columns
- ❖ Data types: integers, strings, dates, floats and null.
- ❖ Columns:
  - ❖ hotel, is\_canceled,
  - ❖ lead\_time, arrival\_date...,
  - ❖ meal, children, country,
  - ❖ number of booking changes,
  - ❖ customer type, reserved room type,
  - ❖ parking space, reservation status etc



# DATASET CLEANING

- ❖ 129,425 null values where detected and replaced.
- ❖ 31,994 duplicate rows where detected and removed.

● Null values      ● Duplicate rows



# EXPLORATORY ANALYSIS



HOTEL

YEAR & MONTH

COUNTRY

MEAL

DISTRIBUTION CHANNEL

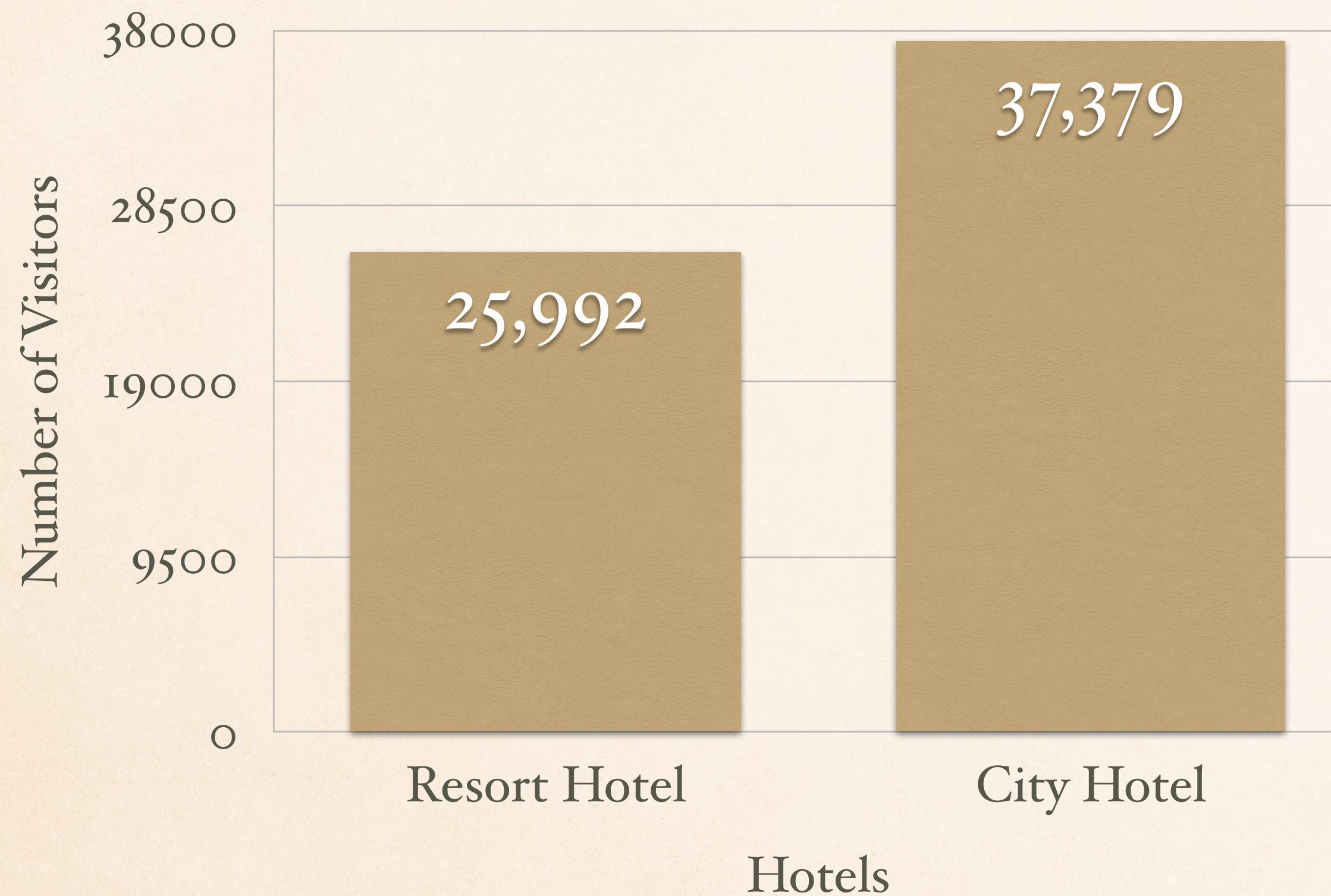
CUSTOMER TYPE

RESERVED ROOM TYPE

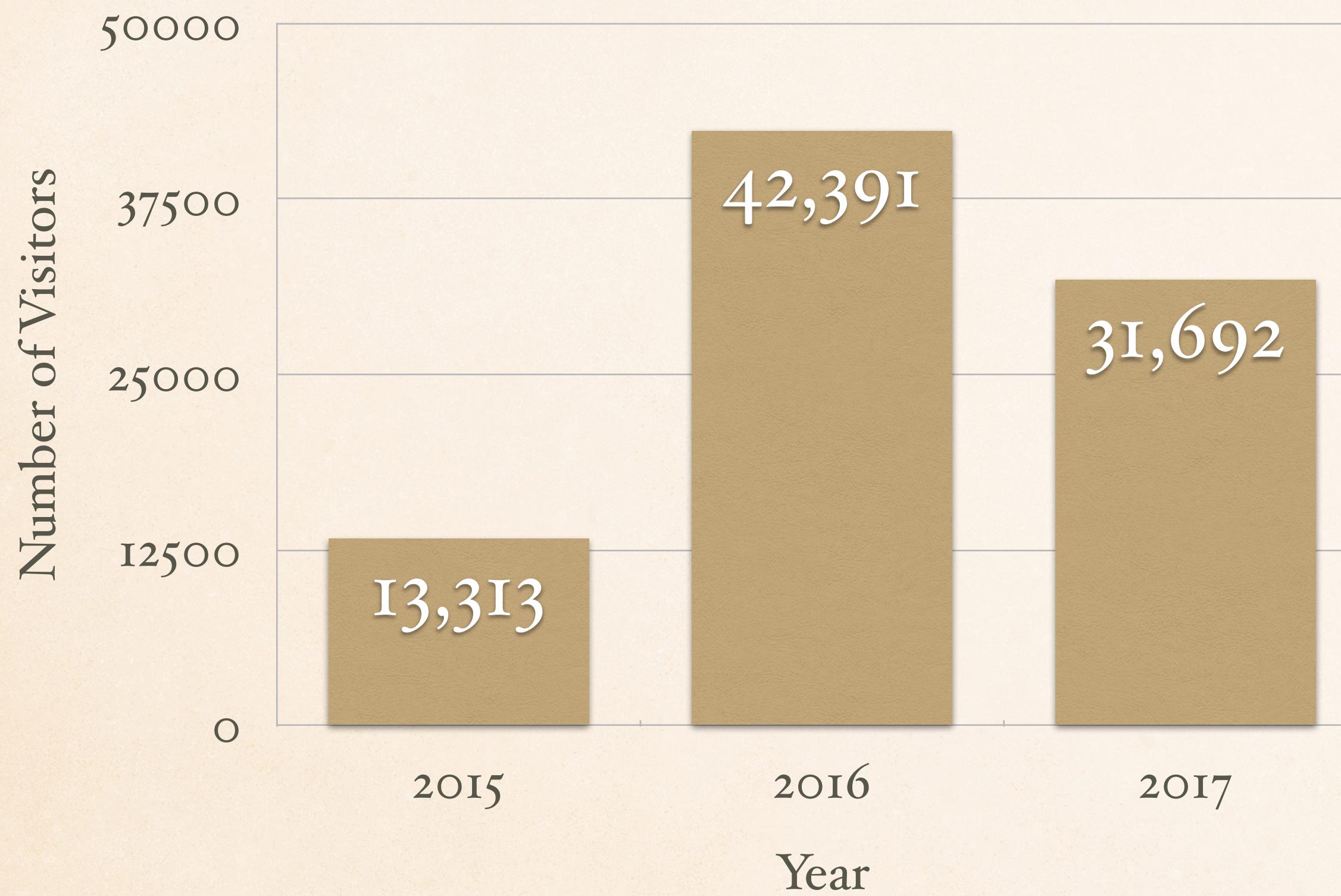
PARKING SPACES



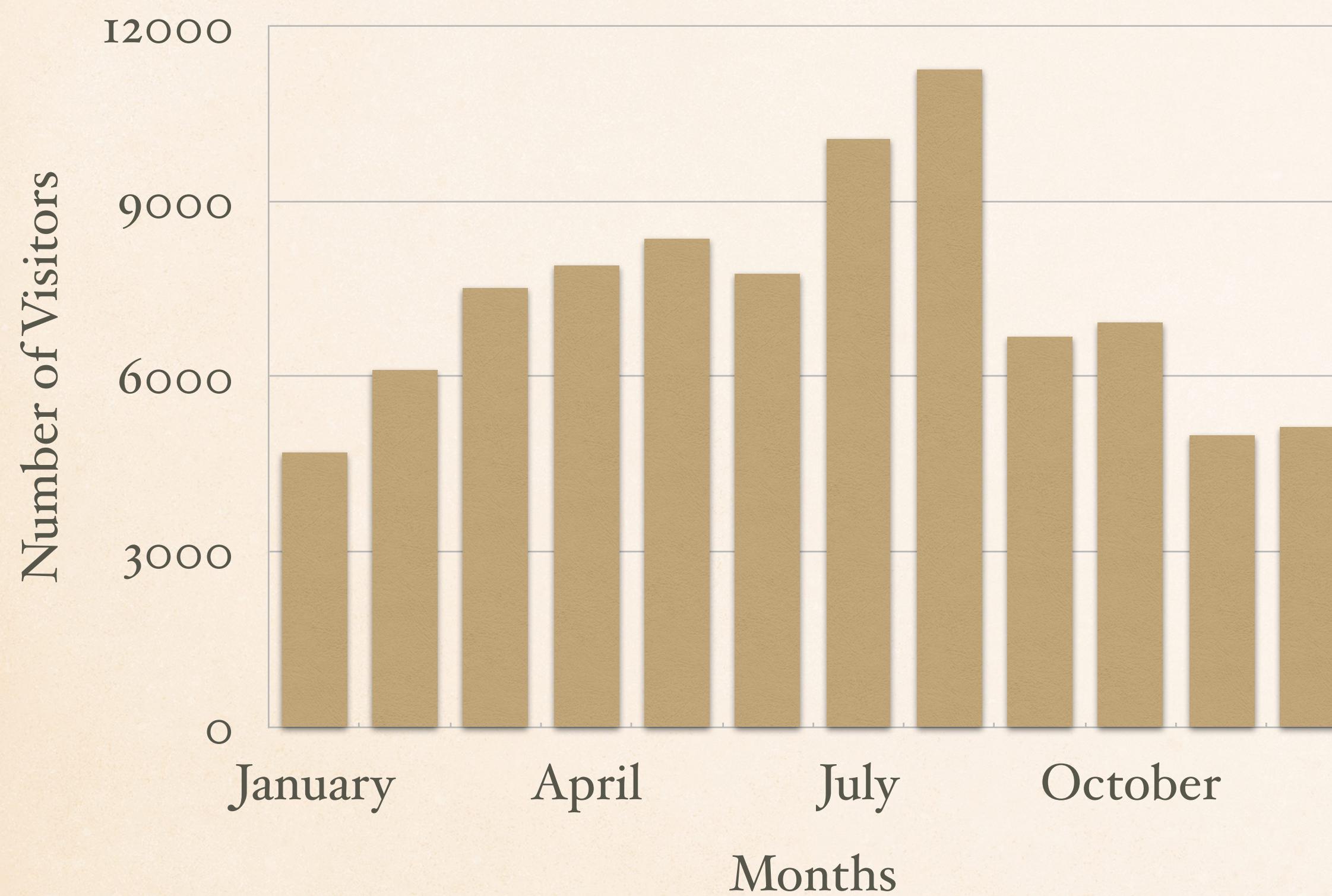
# EXPLORATORY ANALYSIS: HOTEL



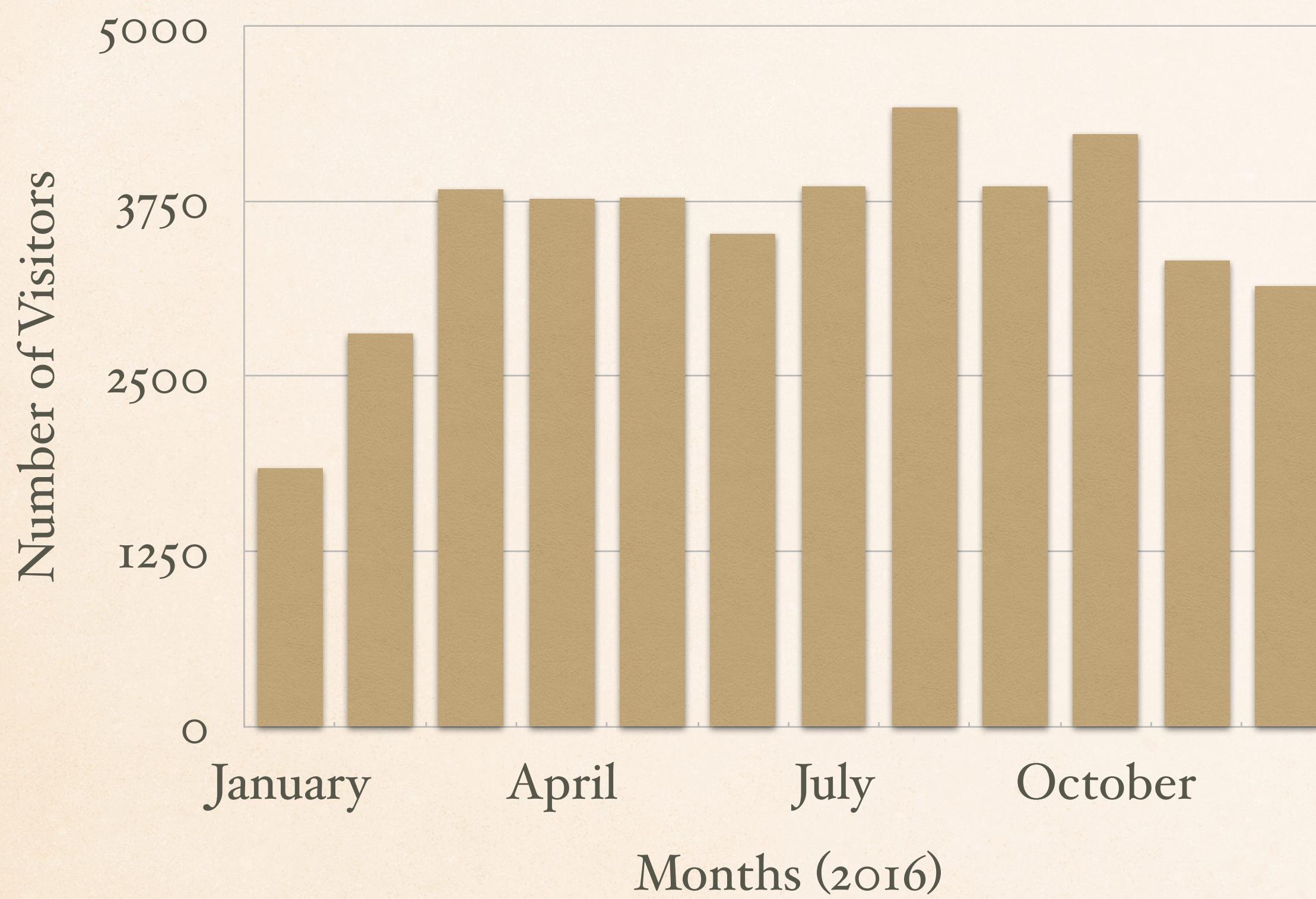
# EXPLORATORY ANALYSIS: YEAR AND MONTH



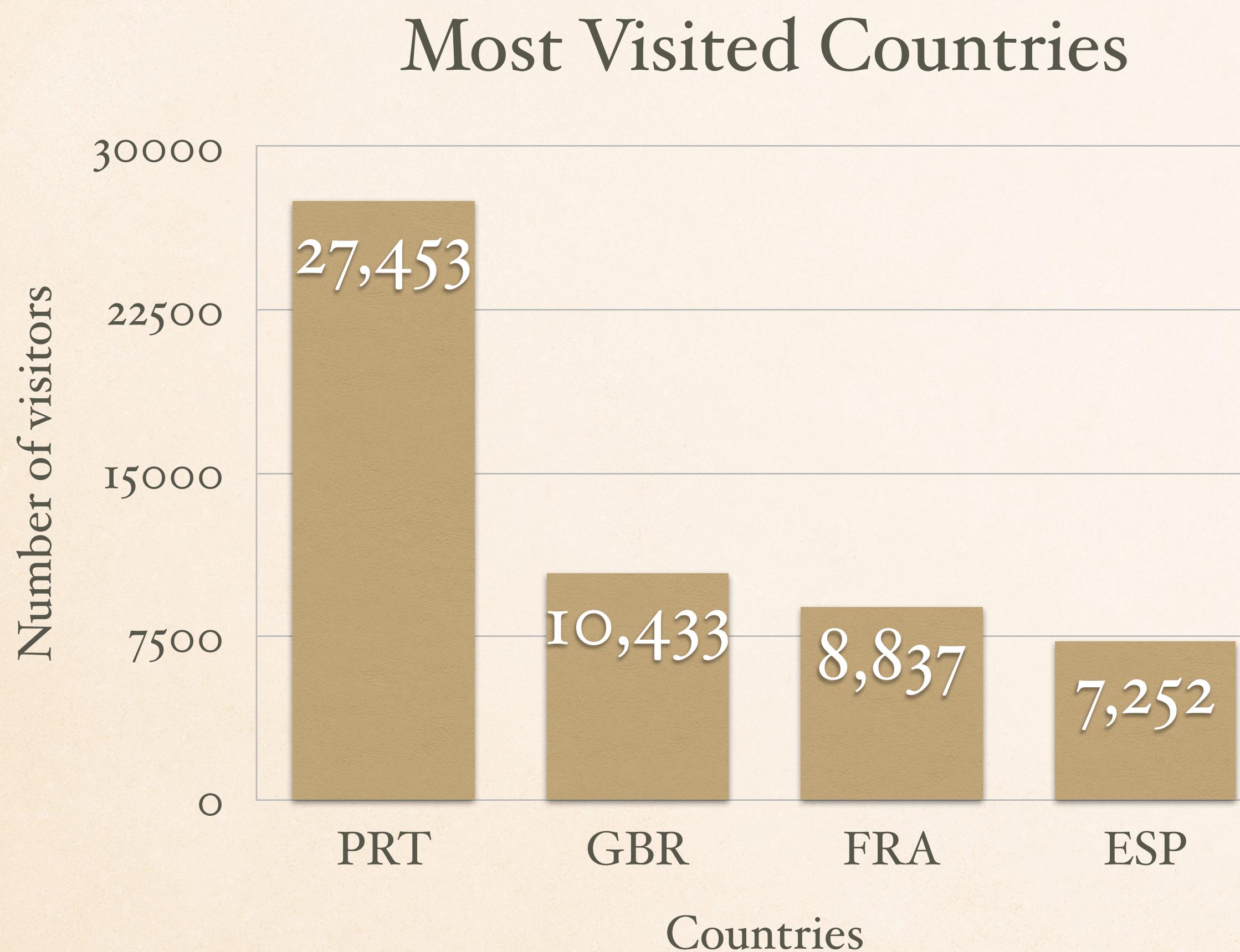
# EXPLORATORY ANALYSIS: YEAR AND MONTH



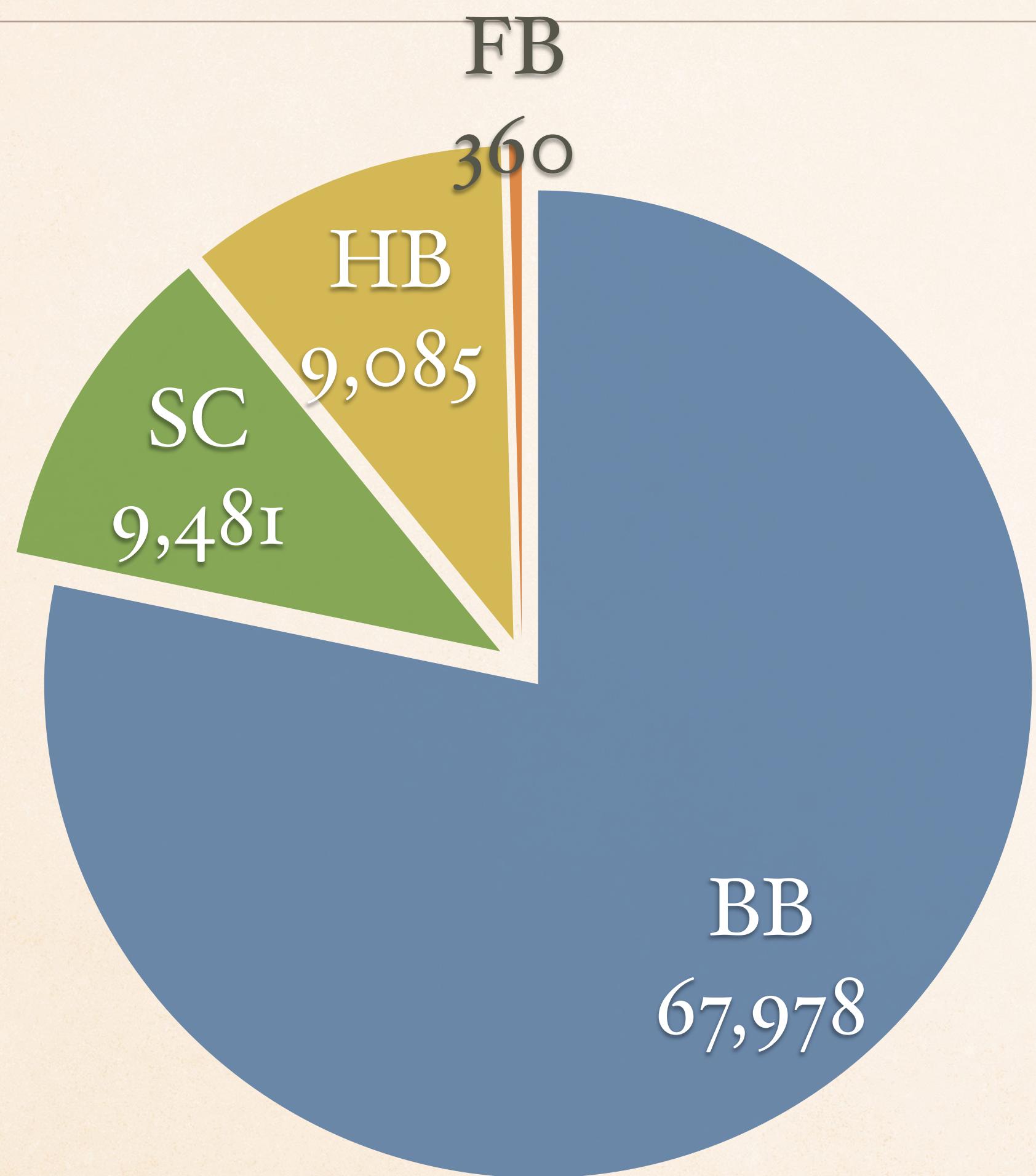
# EXPLORATORY ANALYSIS: YEAR AND MONTH



# EXPLORATORY ANALYSIS: COUNTRY



# EXPLORATORY ANALYSIS: MEAL



# EXPLORATORY ANALYSIS: CUSTOMER TYPE

Contract

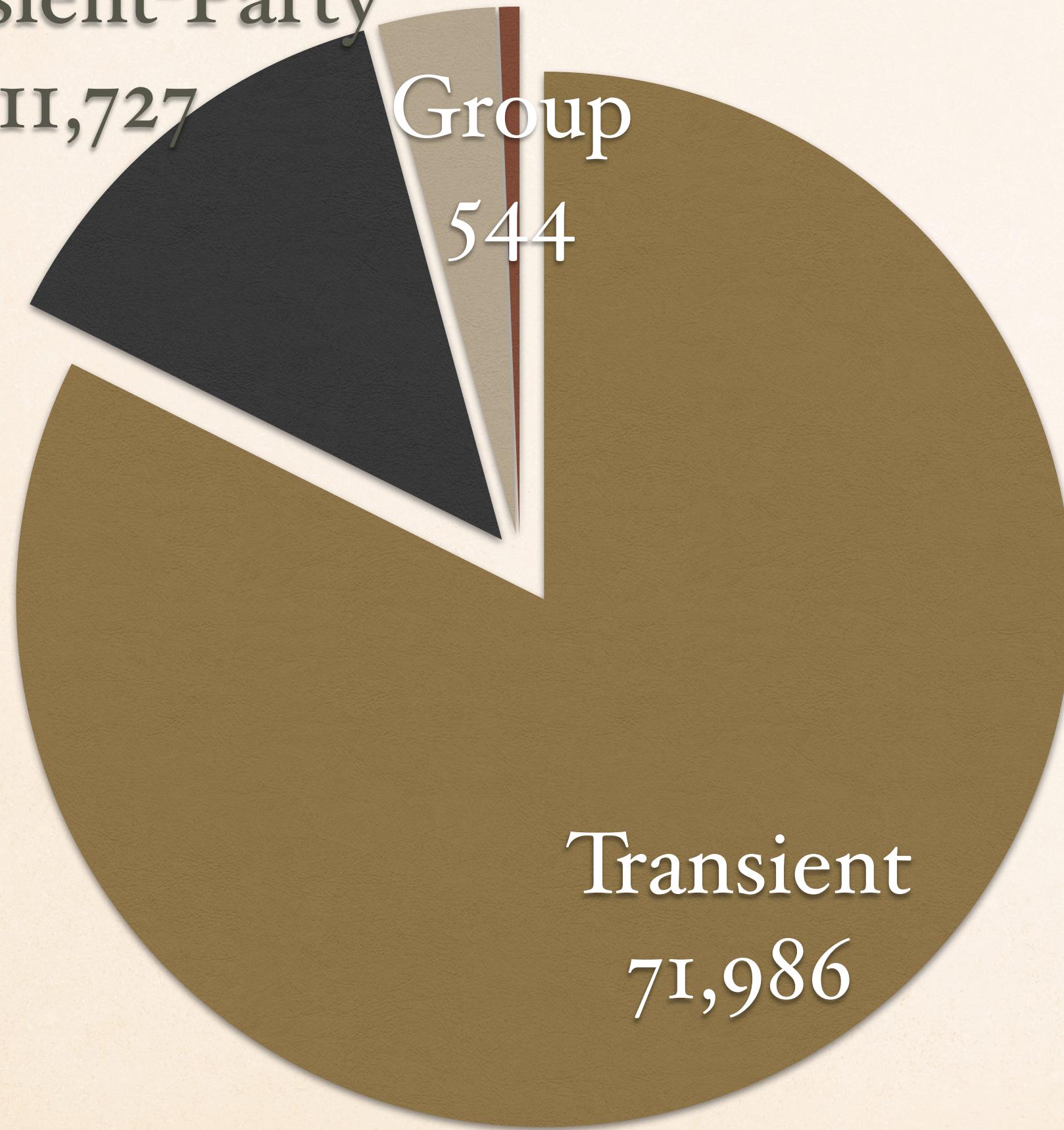
3,139

Transient-Party

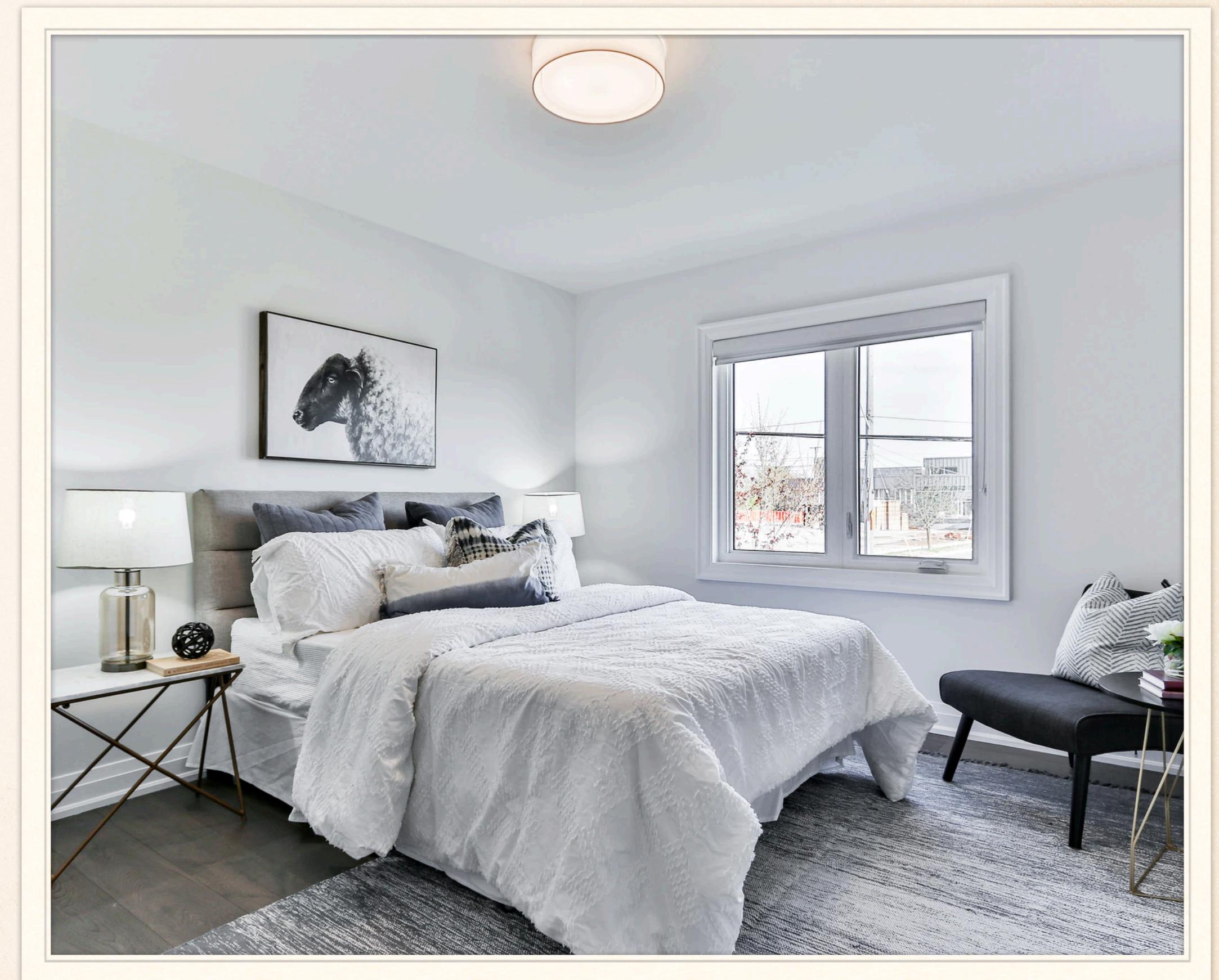
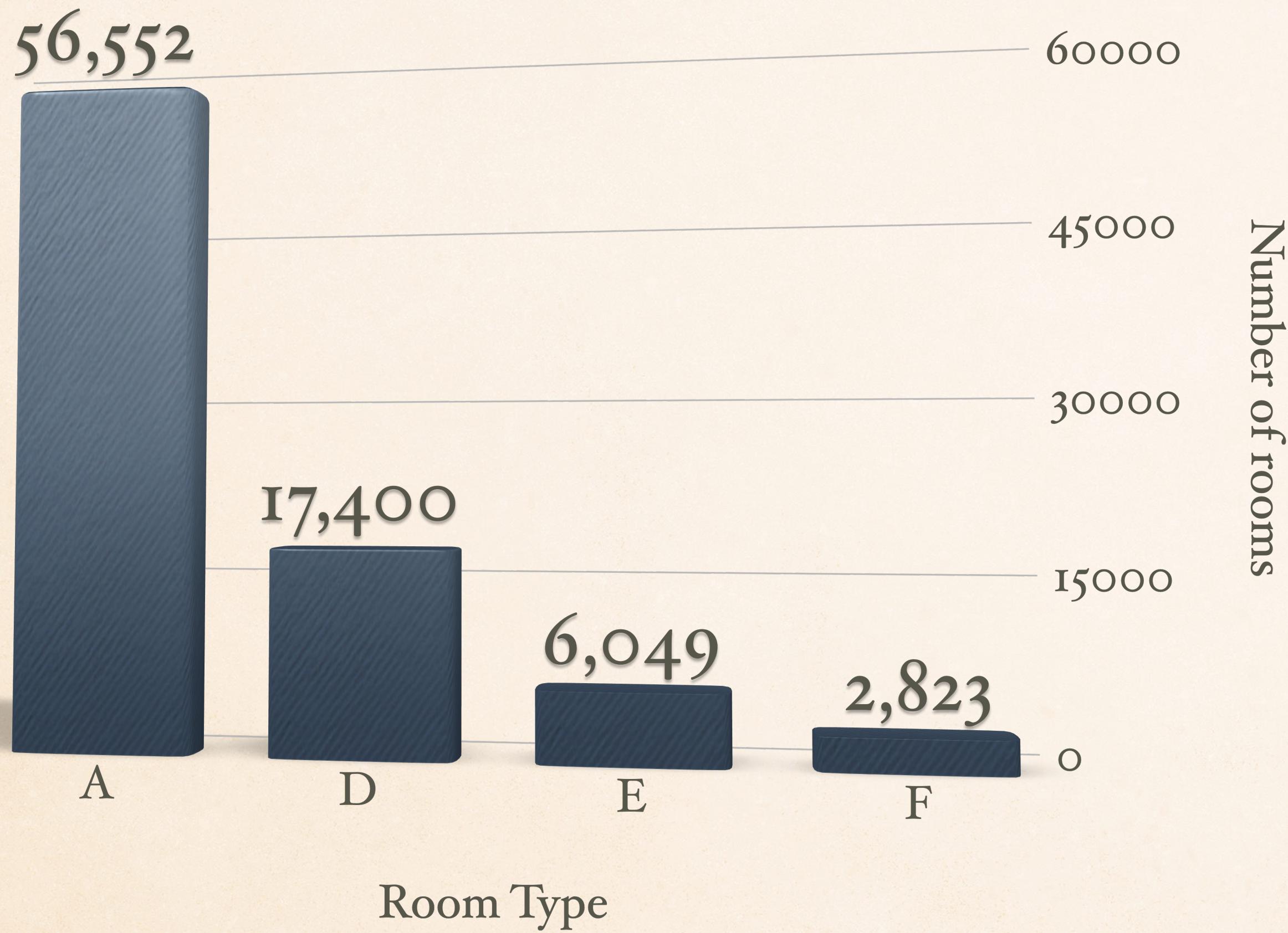
11,727

Group  
544

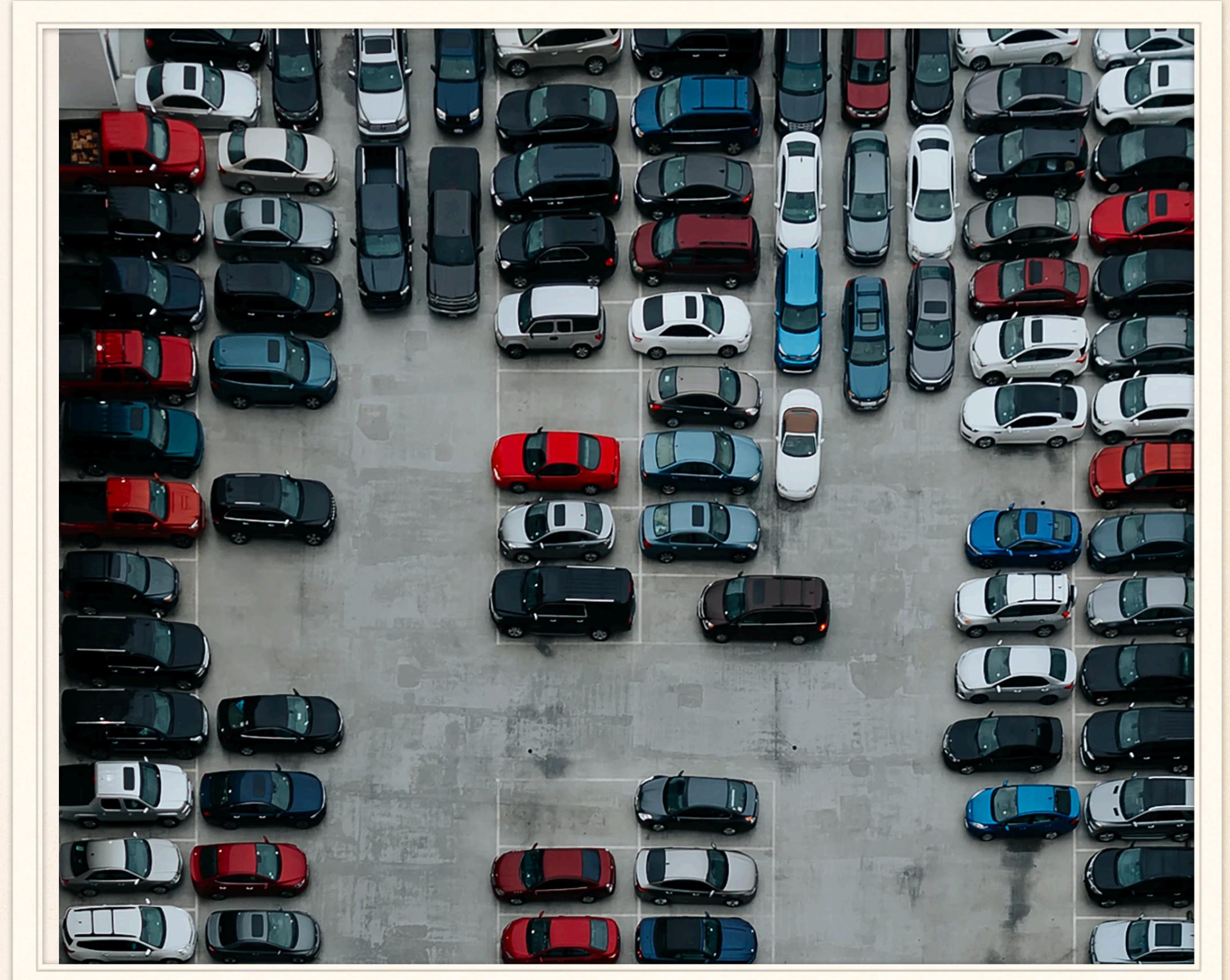
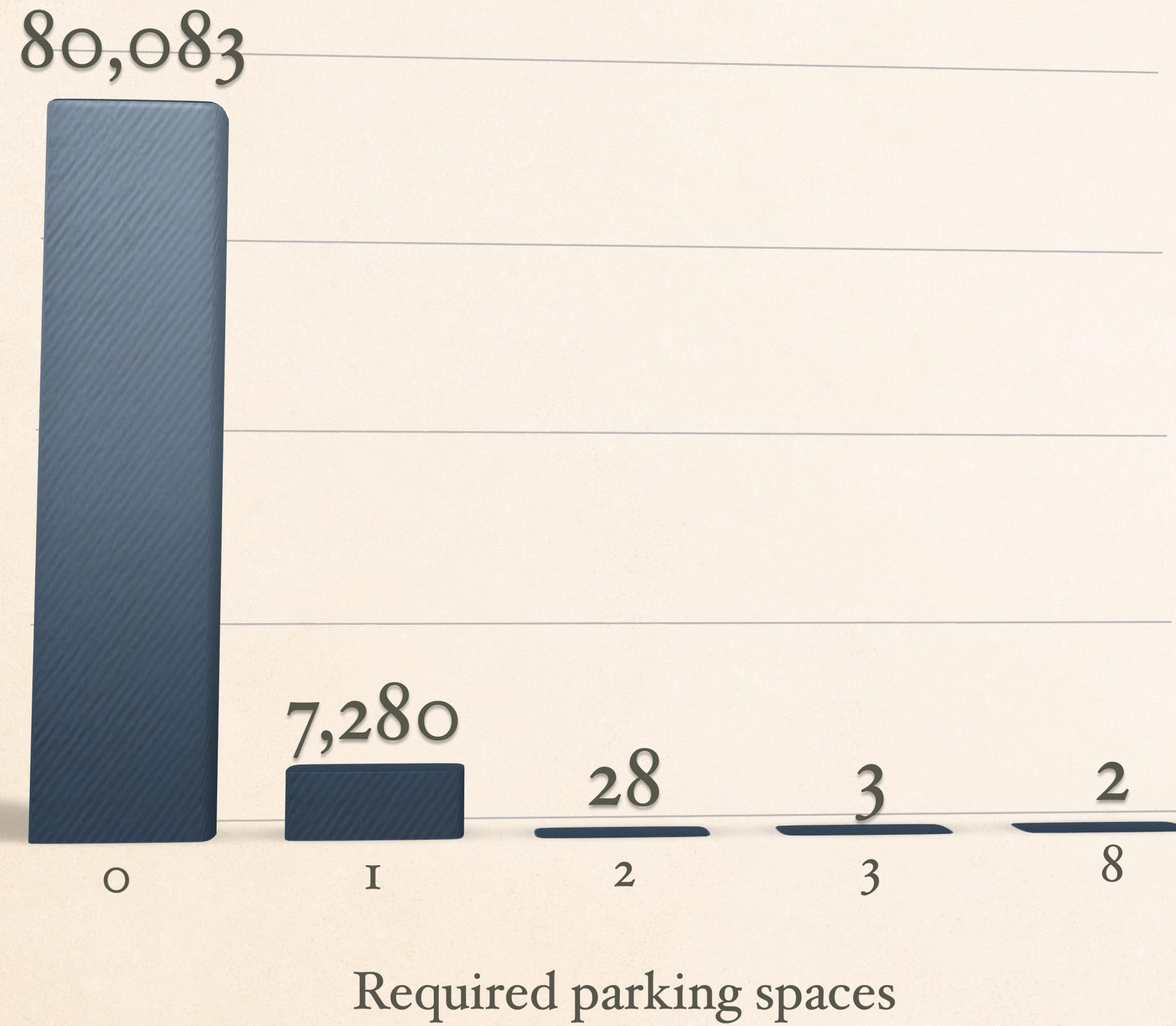
Transient  
71,986



# EXPLORATORY ANALYSIS: RESERVED ROOM



# EXPLORATORY ANALYSIS: PARKING SPACES



# HYPOTHESES

---

HYPOTHESIS I

HYPOTHESIS II



# HYPOTHESIS I

Portugal is the most visited country

because:

- ❖ The temperature, [1-3]
- ❖ Proximity to the sea..., beaches,
- ❖ Hotel cost,
- ❖ Cost of living,
- ❖ Population density,
- ❖ Tourism policy, etc [4]



# HYPOTHESIS II

Majority of the reservation include breakfast because:

- ❖ Breakfast is the most important meal of the day [5-6]



# SEASONALITY

---

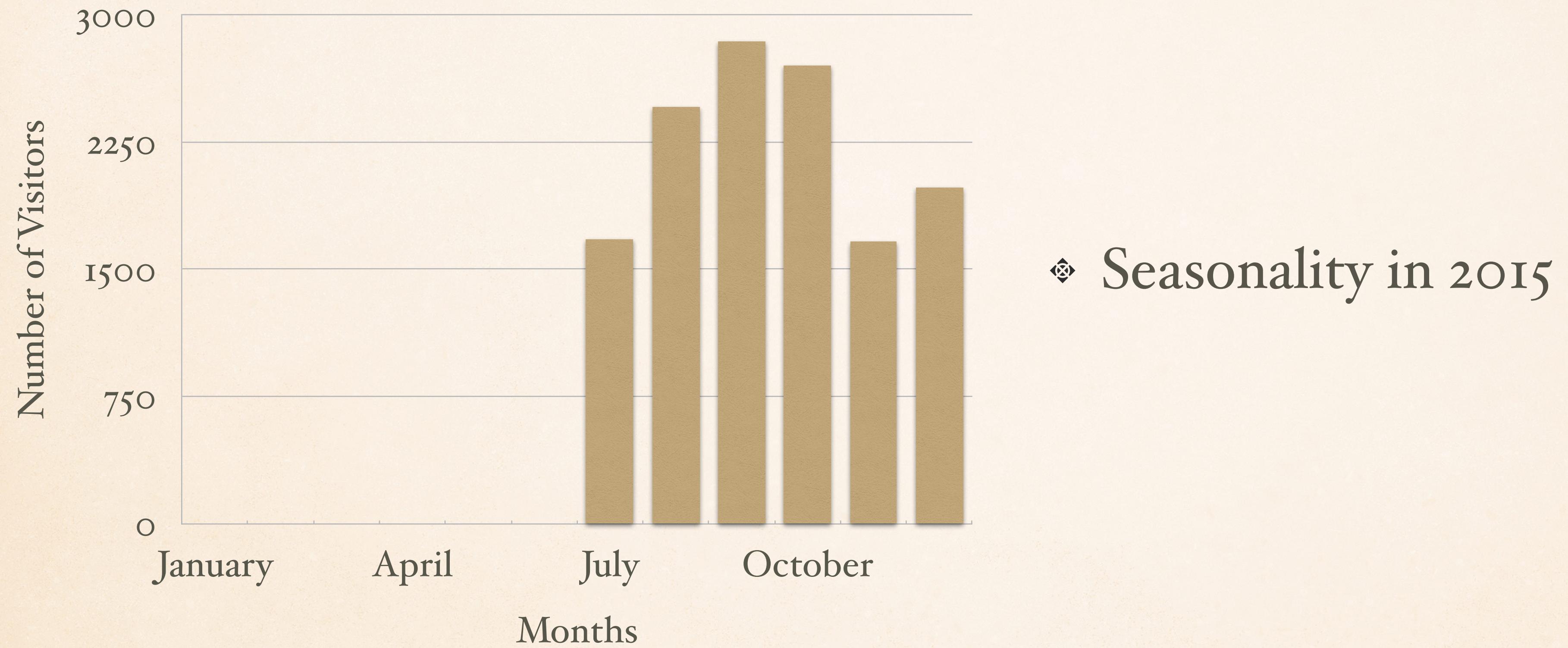


# SEASONALITY

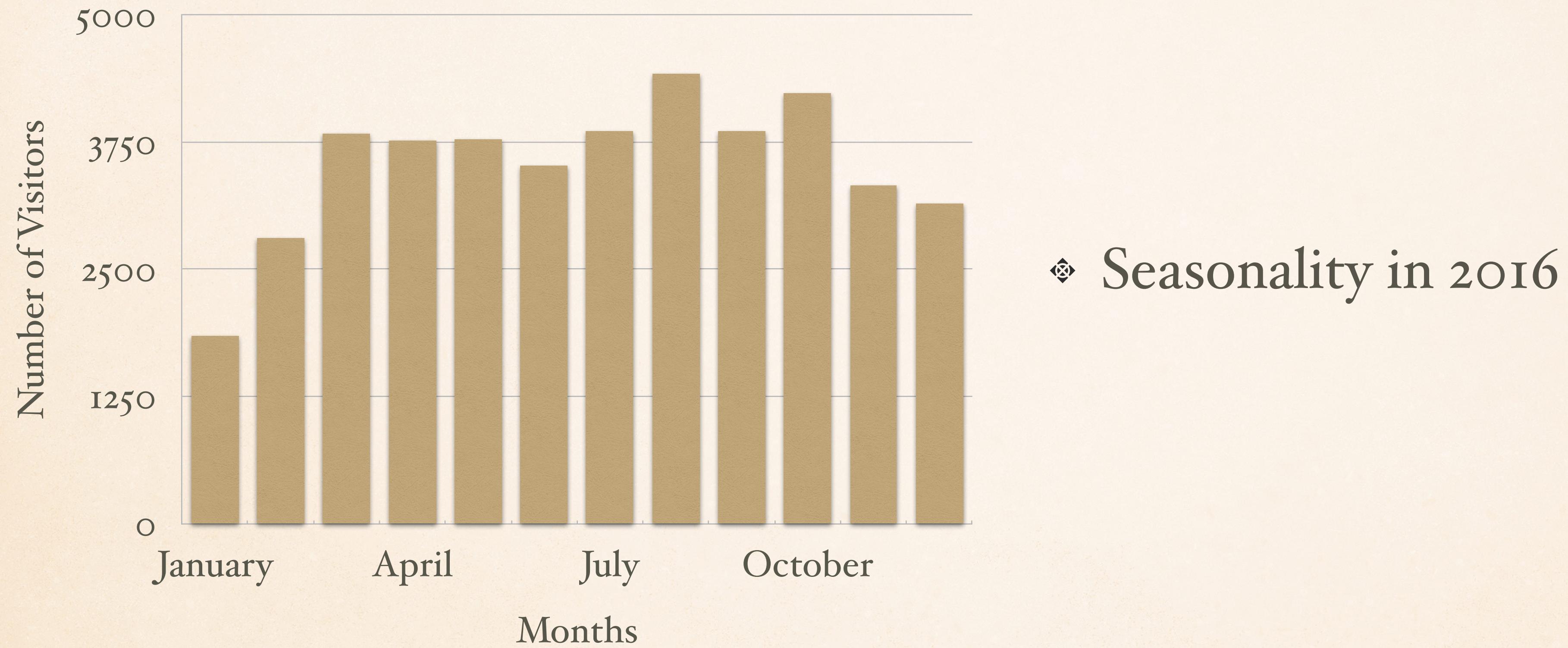
- ❖ Hypothesis I already indicates seasonality, why people travel to Portugal.
- ❖ Historically, there are more reservations/travels/touristic activities in the summer than any other season.



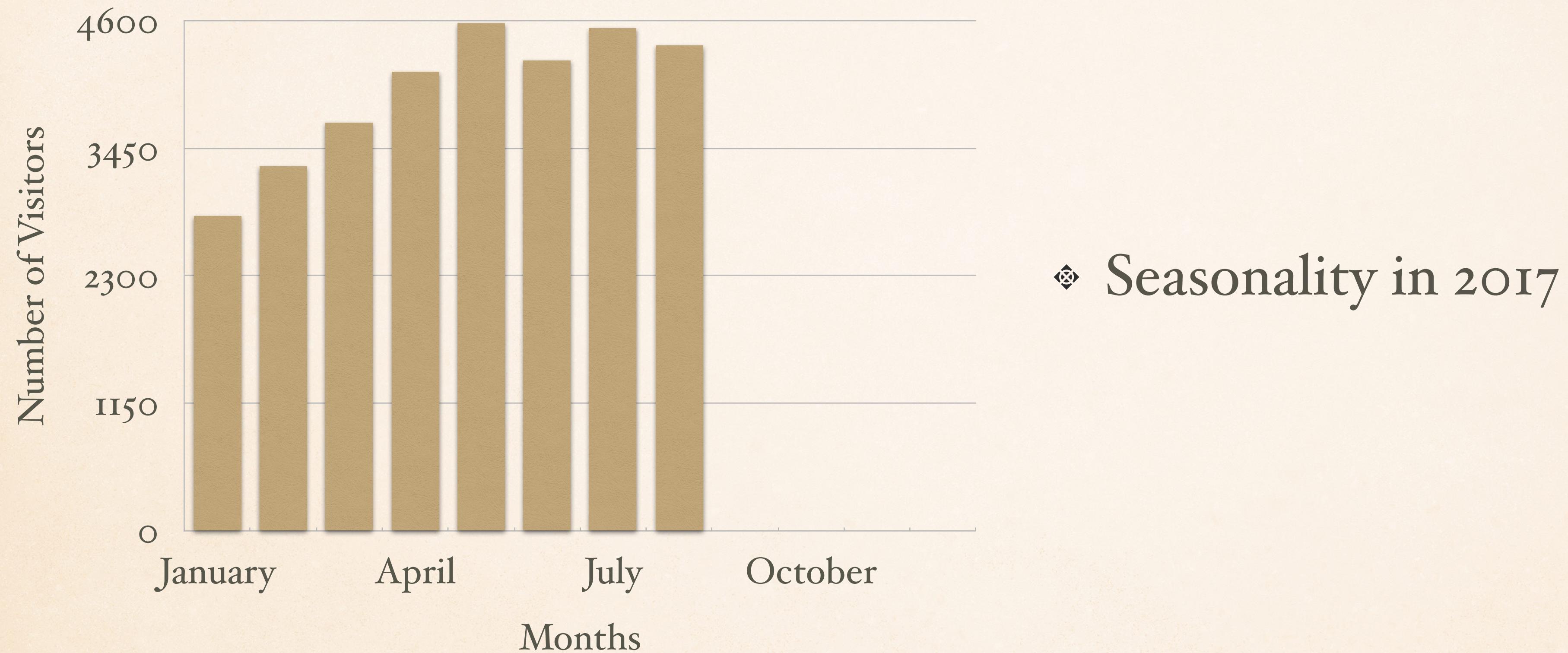
# SEASONALITY



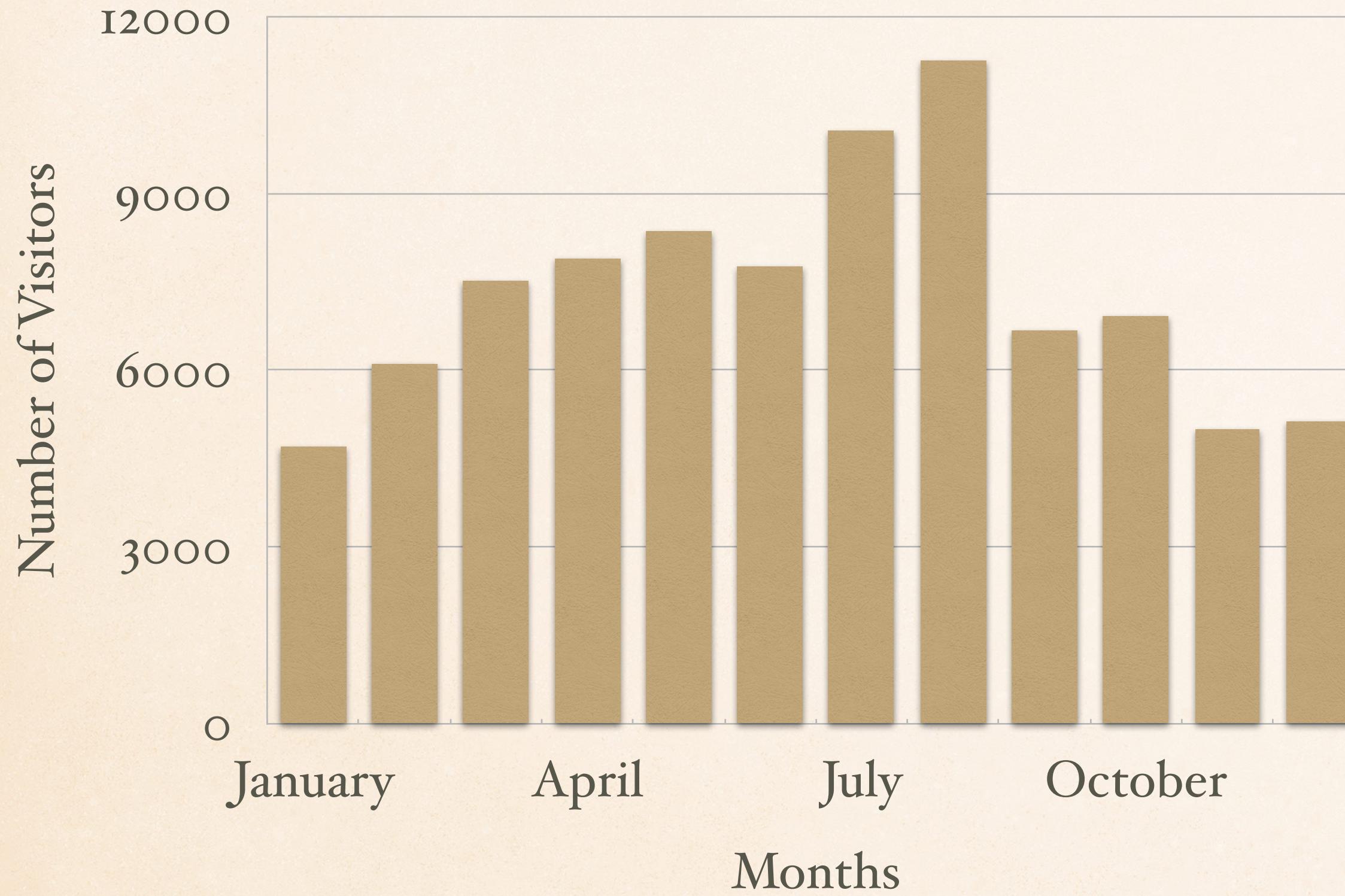
# SEASONALITY



# SEASONALITY



# SEASONALITY



- ❖ Autumn & Winter had the least reservation.
- ❖ Reservations started increasing in the spring as the temperature gets warmer.

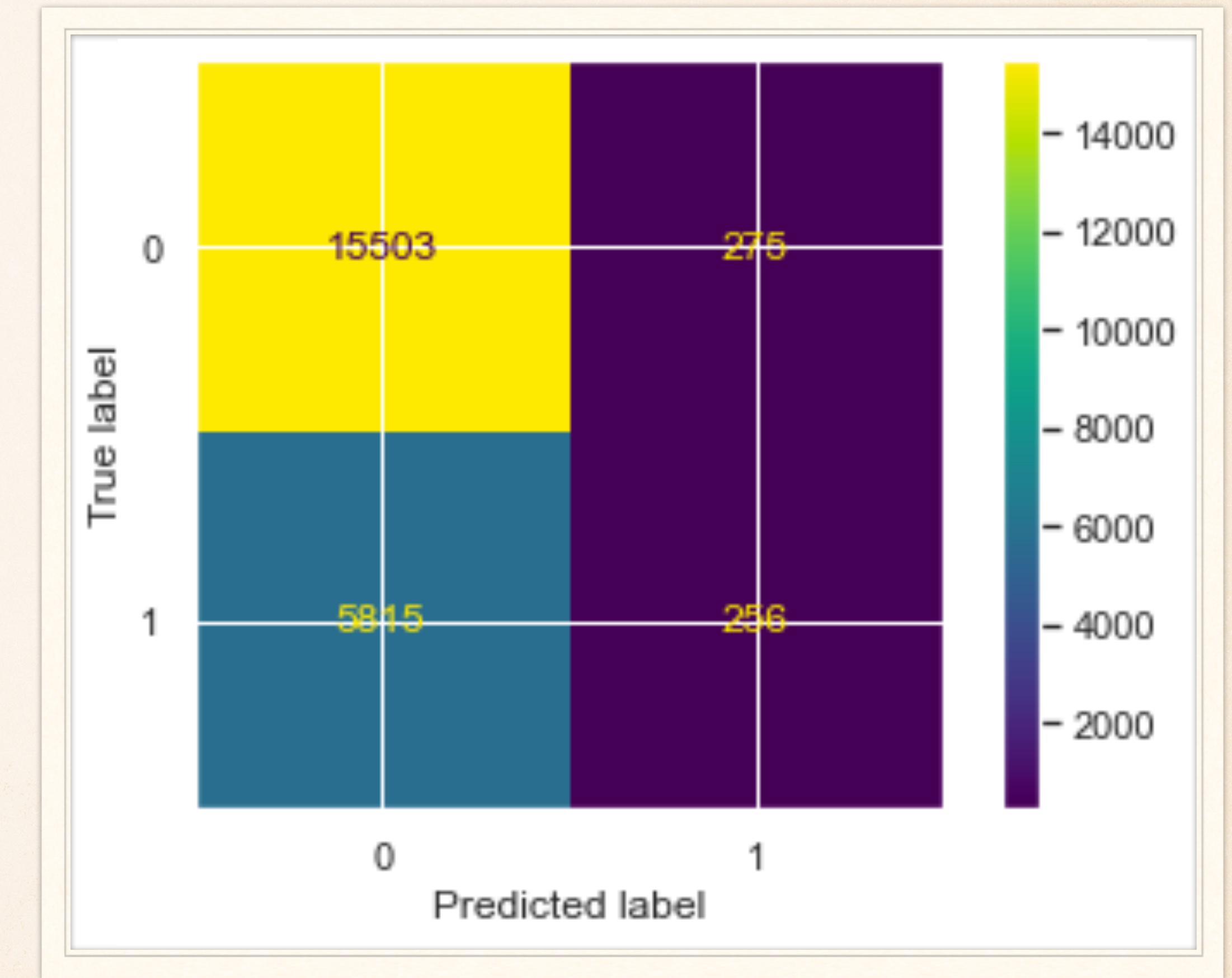
# LEAD TIME VS CANCELLATION

---



## LEAD TIME VS CANCELLATION

- ❖ Built ML model that predicts whether a customer their will cancel reservation given the lead time. (Accuracy = 0.72)
- ❖ The higher the lead time the more probable that the customer will cancel their booking.
- ❖ This explains why some hotels have restrictions or some insurance policy.



“Thank you for your audience.”

*-Yousele Michael*

# REFERENCES

1. Holiday Weather, Algarve, Portugal: Annual Weather Averages. <https://www.holiday-weather.com/algarve/averages/>
2. Holiday Weather, London, United Kingdom: Annual Weather Averages. <https://www.holiday-weather.com/london/averages/>
3. The Portugal News, Portugal's €7bn climate change bill, 2017. <https://www.theportugalnews.com/news/portugals-7bn-climate-change-bill/40860#:~:text=In%20terms%20of%20the%20effects,by%20temperatures%20reached%20in%201989>
4. Almeida García, Fernando & Macias, Rafael & Balbuena, Antonia. (2015). Tourist Spaces and Tourism Policy in Spain and Portugal. International Journal of Culture Tourism and Hospitality Research. 10. 235-249. 10.1108/S1871-317320150000010017.
5. Charles Spence, Breakfast: The most important meal of the day?, International Journal of Gastronomy and Food Science, Volume 8, 2017, Pages 1-6, ISSN 1878-450X, <https://doi.org/10.1016/j.ijgfs.2017.01.003>.
6. Benton et al., 2001 D. Benton, O. Slater, R.T. Donohoe The influence of breakfast and a snack on psychological functioning Physiol. Behav., 74 (2001), pp. 559-571