

# Topic ideas

LAM-duh: Xuliang Deng, Leah Okamura, Megan Richards

Update date

## Data Set 1 - XULIANG

### Introduction and Data

The dataset we are using is from <https://github.com/ayoubimaya/pokemon> and is sourced from <https://www.pokemon.com/us/pokedex/>, which is the official Pokemon open database. In this dataset, every observation is a specific pokemon from the pokedex. The general characteristics being measured in the data are the name, type, and characteristics of each Pokemon. More specifically, some variables that we are interested in are Attack, HP, Defense, Speed, and the Pokemon type.

### Research questions

We would like to see if there are any differences in the means of attack, defense, HP, based on each type using an ANOVA test. We are interested in seeing how these means might differ to each other through hypothesis testing. We are also interested in analyzing how variables such as attack, defense, and speed may affect the total points assigned to a specific pokemon.

## Data Set 2 - MEGAN

### Introduction and Data

(Introduce and discuss data here)

### Research questions

(Discuss research questions here)

## Data Set 3 - LEAH

### Introduction and Data

The source of the dataset is Tiny Tuesday, <https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-02-11/readme.md>. This data set comes from an open hotel booking demand dataset from Antonio, Almeida and Nunes, 2019. It is sourced from this study <https://www.sciencedirect.com/science/article/pii/S2352340918315191#f0010>. Due to the dataset being over 100,000 observations, we have limited

the observations to be only hotels from the US. The general characteristics being measured in the data are the different aspects of booking and staying at a hotel. For example, out of the 32 variables, some of the ones we find great interest in are hotel type, reserved room type, assigned room type, company, meal, number of adults/children/babies, the average daily rate or daily cost, and the reservation status.

## Research questions

A research question we are interested in is how do factors such as type of hotel and type of guest affect the average daily rate for a hotel. We would also be interested in seeing how stays in the weekend or the weekday may affect the average daily rate for a hotel, and if they differ between the two hotel types, City and Resort hotels.

## Glimpse of data sets

### Data set 1 - XULIANG

```
## Rows: 800
## Columns: 13
## $ '#<dbl> 1, 2, 3, 3, 4, 5, 6, 6, 6, 7, 8, 9, 9, 10, 11, 12, 13, 14, ~
## $ Name <chr> "Bulbasaur", "Ivysaur", "Venusaur", "VenusaurMega Venusaur"~
## $ 'Type 1' <chr> "Grass", "Grass", "Grass", "Grass", "Fire", "Fire", "Fire",~
## $ 'Type 2' <chr> "Poison", "Poison", "Poison", "Poison", NA, NA, "Flying", "~
## $ Total <dbl> 318, 405, 525, 625, 309, 405, 534, 634, 634, 314, 405, 530,~
## $ HP <dbl> 45, 60, 80, 80, 39, 58, 78, 78, 78, 44, 59, 79, 79, 45, 50,~
## $ Attack <dbl> 49, 62, 82, 100, 52, 64, 84, 130, 104, 48, 63, 83, 103, 30,~
## $ Defense <dbl> 49, 63, 83, 123, 43, 58, 78, 111, 78, 65, 80, 100, 120, 35,~
## $ 'Sp. Atk' <dbl> 65, 80, 100, 122, 60, 80, 109, 130, 159, 50, 65, 85, 135, 2~
## $ 'Sp. Def' <dbl> 65, 80, 100, 120, 50, 65, 85, 85, 115, 64, 80, 105, 115, 20~
## $ Speed <dbl> 45, 60, 80, 80, 65, 80, 100, 100, 100, 43, 58, 78, 78, 45, ~
## $ Generation <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ Legendary <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FAL~
```

### Data set 2 - MEGAN

### Data set 3 - LEAH

```
## Rows: 2,097
## Columns: 32
## $ hotel <chr> "Resort Hotel", "Resort Hotel", "Resort~
## $ is_canceled <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ lead_time <dbl> 68, 14, 10, 9, 51, 51, 98, 88, 10, 42, ~
## $ arrival_date_year <dbl> 2015, 2015, 2015, 2015, 2015, 2015, 201~
## $ arrival_date_month <chr> "July", "July", "July", "July", "July",~
## $ arrival_date_week_number <dbl> 27, 27, 27, 27, 28, 28, 28, 28, 28, 29,~
## $ arrival_date_day_of_month <dbl> 1, 2, 3, 3, 6, 6, 6, 7, 10, 13, 16, 28,~
## $ stays_in_weekend_nights <dbl> 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 2, ~
## $ stays_in_week_nights <dbl> 4, 2, 2, 1, 3, 3, 1, 4, 1, 2, 1, 1, 8, ~
## $ adults <dbl> 2, 2, 2, 2, 2, 3, 2, 3, 2, 2, 2, 2, 2, ~
## $ children <dbl> 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 2, 1, 0, ~
## $ babies <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
```

```

## $ meal <chr> "BB", "BB", "BB", "BB", "BB", "BB", "BB~
## $ country <chr> "USA", "USA", "USA", "USA", "USA", "USA~
## $ market_segment <chr> "Online TA", "Online TA", "Online TA", ~
## $ distribution_channel <chr> "TA/TO", "TA/TO", "TA/TO", "TA/TO", "TA~
## $ is_repeated_guest <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ previous_cancellations <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ previous_bookings_not_canceled <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ reserved_room_type <chr> "D", "A", "G", "C", "G", "G", "D", "D", ~
## $ assigned_room_type <chr> "E", "C", "H", "C", "G", "G", "F", "E", ~
## $ booking_changes <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, ~
## $ deposit_type <chr> "No Deposit", "No Deposit", "No Deposit~
## $ agent <chr> "240", "242", "240", "241", "241", "241~
## $ company <chr> "NULL", "NULL", "NULL", "NULL", "NULL", ~
## $ days_in_waiting_list <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ customer_type <chr> "Transient", "Transient", "Transient", ~
## $ adr <dbl> 97.00, 98.00, 153.00, 94.71, 117.81, 11~
## $ required_car_parking_spaces <dbl> 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, ~
## $ total_of_special_requests <dbl> 3, 1, 0, 0, 2, 2, 1, 1, 0, 0, 1, 1, 1, ~
## $ reservation_status <chr> "Check-Out", "Check-Out", "Check-Out", ~
## $ reservation_status_date <chr> "7/5/15", "7/4/15", "7/5/15", "7/4/15", ~

```