

Analysis Report

MG

Modified: 2023-05-13

Contents

1 Business Task Statement	1
2 Deliverables	1
3 Data	1
3.1 mg_resorts.csv and mg_resorts_gs.csv	2
3.2 Data Structure	2
4 Data Processing	5
4.1 Data Cleaning	5
4.2 Missing Values	7
4.3 White Spaces	7
5 Data Credibility	7
6 Data Bias	7
7 License	7
8 Future Work	7
References	8

1 Business Task Statement

The purpose of this study is to build a dashboard to help skiers find their ideal destination for the next winter season. The dashboard should allow for exploration and interactivity.

The original dataset includes details on 499 ski resorts around the world (location, slopes, lifts, prices, and ski season). There is also a table with the snow cover around the world for each month of 2022 by latitude and longitude. It contains two data tables in CSV format and description of fields.

2 Deliverables

The dashboard in Shiny.

3 Data

The original data is located at [1]. The data is organized in two .csv files:

- mg_data_dictionary.csv - the description of fields.

- `mg_resorts.csv` - the actual data (original data)
- `mg_resorts_gs.csv` - the processed file
- `mg_snow.csv` - snow cover for selected GPS locations for the entire 2022 year.

It is not specified which year the resorts data regards. However, I expect that it covers 2022 year. In addition, [1] states that the data source is <https://ski-resort-stats.com> and NASA Earth Observations [2] with License *Public Domain*.

3.1 `mg_resorts.csv` and `mg_resorts_gs.csv`

File	Data Structure	No. of Fields	No. of Records	Date Added
<code>mg_resorts.csv</code>	Table	25	499	02/07/2023
<code>mg_resorts_gs.csv</code>	Table	42	480	02/07/2023

After data examination, it occurred that there are duplicates (explained later). I also reorganized fields for seasons, splitted seasons to months, added the field ‘Seasons number’ indicating the number of seasons shown in the original file. I added the numerical columns corresponding to last four fields with ‘Yes’ ‘No’ answers (there were no other answers).

For analysis I use the input/`mg_resorts_gs.csv` file.

3.2 Data Structure

The description of main fields is in `mg_data_dictionary.csv`:

For the input/`mg_resorts_gs.csv` file.

Name	Description	Notes
ID	Identifier for each resort	
Resort	Name of the ski and snowboard resort	
Country	Resort country location	Unique countries: 38
Continent	Resort continent location	Unique continents: 5
Price	Ski pass cost for 1 adult for 1 day in the main season in Euro	Season
Months	0 indicates a resort is closed, 1 the resort is open	
Highest point	Highest mountain point in meters	
Lowest point	Lowest possible point to ski in meters	
Beginner slopes	Total length of “children”, “blue”, and “green” slopes in km	
Intermediate slopes	Total length of “red” slopes in km	
Difficult slopes	Total length of “black”, “advanced”, and “expert” slopes in km	
Total slopes	Total length of slopes in km	
Longest run	Longest possible continuous run in km	
Snow cannons	Total amount of snow cannons	

Name	Description	Notes
Surface lifts	Total number of surface lifts, including T-bar, Sunkidslift, Rope lifts and people mover	
Chair lifts	Total number of chair lifts	
Gondola lifts	Total number of gondola lifts, including Gondola, Train lifts, Funicular, Combined gondola and chairlifts, Helicopter lifts, Snowcats and Aerial tramways	
Total lifts	Total number of lifts	
Lift capacity	Number of passengers the resort's lift system can move in an hour	
Child friendly	Is the ski resort child friendly?	Categorical
Org Child friendly	Is the ski resort child friendly?	0 - means 'no', 1 - means 'yes'
Snowparks Org	Does the resort have one or more snowparks?	Categorical
Snowparks	Does the resort have one or more snowparks?	0 - means 'no', 1 - means 'yes'
Nightskiing Org	Does the resort offer skiing on illuminated slopes?	0 - means 'no', 1 - means 'yes'
Summer skiing Org	Does the resort offer skiing during the summer?	Categorical
Summer skiing	Does the resort offer skiing during the summer?	0 - means 'no', 1 - means 'yes'

Processed file input/mg_resorts_gs.csv summary

```
##          ID          Resort          Latitude          Longitude
##  Min.    : 1.0    Length:480    Min.    :-45.05    Min.    :-149.7407
##  1st Qu.:131.8    Class :character    1st Qu.: 43.32    1st Qu.:  0.4226
##  Median :259.5    Mode  :character    Median : 46.36    Median :  8.4056
##  Mean   :255.6                      Mean   : 43.09    Mean   : -6.5730
##  3rd Qu.:379.2                      3rd Qu.: 47.35    3rd Qu.: 12.5641
##  Max.   :499.0                      Max.   : 67.78    Max.   : 176.8767
##  Country          Continent          Price          Season
##  Length:480          Length:480          Min.    : 0.00    Length:480
##  Class :character    Class :character    1st Qu.: 35.00    Class :character
##  Mode  :character    Mode  :character    Median : 45.00    Mode  :character
##                      Mean   : 48.39
##                      3rd Qu.: 54.00
##                      Max.   :141.00
##  January          February          March          April
##  Min.    :0.0000    Min.    :0.0000    Min.    :0.0000    Min.    :0.0000
##  1st Qu.:1.0000    1st Qu.:1.0000    1st Qu.:1.0000    1st Qu.:1.0000
##  Median :1.0000    Median :1.0000    Median :1.0000    Median :1.0000
##  Mean   :0.9958    Mean   :0.8833    Mean   :0.8875    Mean   :0.8083
##  3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.:1.0000
```

##	Max. :2.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000
##	May	June	July	August
##	Min. :0.0000	Min. :0.00000	Min. :0.00000	Min. :0.00000
##	1st Qu.:0.0000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000
##	Median :0.0000	Median :0.00000	Median :0.00000	Median :0.00000
##	Mean :0.1271	Mean :0.06042	Mean :0.05833	Mean :0.05625
##	3rd Qu.:0.0000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :1.0000	Max. :1.00000	Max. :1.00000	Max. :1.00000
##	September	October	November	December
##	Min. :0.00000	Min. :0.00000	Min. :0.0000	Min. :0.0000
##	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.0000	1st Qu.:1.0000
##	Median :0.00000	Median :0.00000	Median :0.0000	Median :1.0000
##	Mean :0.05833	Mean :0.08125	Mean :0.3479	Mean :0.8875
##	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:1.0000	3rd Qu.:1.0000
##	Max. :1.00000	Max. :1.00000	Max. :1.0000	Max. :1.0000
##	Seasons Count	Highest point	Lowest point	Beginner slopes
##	Min. :0.00	Min. : 163	Min. : 36.0	Min. : 0.00
##	1st Qu.:1.00	1st Qu.:1558	1st Qu.: 788.2	1st Qu.: 9.00
##	Median :1.00	Median :2122	Median :1127.5	Median : 18.00
##	Mean :0.95	Mean :2131	Mean :1205.1	Mean : 26.07
##	3rd Qu.:1.00	3rd Qu.:2640	3rd Qu.:1505.8	3rd Qu.: 30.00
##	Max. :3.00	Max. :3914	Max. :3286.0	Max. :312.00
##	Intermediate slopes	Difficult slopes	Total slopes	Longest run
##	Min. : 0.00	Min. : 0.00	Min. : 1.00	Min. : 0.000
##	1st Qu.: 11.00	1st Qu.: 3.00	1st Qu.: 29.00	1st Qu.: 0.000
##	Median : 24.00	Median : 8.00	Median : 52.00	Median : 3.000
##	Mean : 32.28	Mean : 14.31	Mean : 72.66	Mean : 3.623
##	3rd Qu.: 40.25	3rd Qu.: 20.00	3rd Qu.: 97.75	3rd Qu.: 6.000
##	Max. :239.00	Max. :126.00	Max. :600.00	Max. :16.000
##	Snow cannons	Surface lifts	Chair lifts	Gondola lifts
##	Min. : 0.0	Min. : 0.000	Min. : 0.000	Min. : 0.000
##	1st Qu.: 0.0	1st Qu.: 3.000	1st Qu.: 3.000	1st Qu.: 0.000
##	Median : 11.5	Median : 7.000	Median : 6.000	Median : 1.000
##	Mean : 145.3	Mean : 9.771	Mean : 8.419	Mean : 2.583
##	3rd Qu.: 152.0	3rd Qu.:12.000	3rd Qu.:11.000	3rd Qu.: 3.000
##	Max. :2383.0	Max. :89.000	Max. :74.000	Max. :40.000
##	Total lifts	Lift capacity	Child friendly Org	Child friendly
##	Min. : 0.00	Min. : 0	Length:480	Min. :0.0000
##	1st Qu.: 10.00	1st Qu.: 11200	Class :character	1st Qu.:1.0000
##	Median : 15.00	Median : 17837	Mode :character	Median :1.0000
##	Mean : 20.77	Mean : 26629		Mean :0.9917
##	3rd Qu.: 24.00	3rd Qu.: 29990		3rd Qu.:1.0000
##	Max. :174.00	Max. :252280		Max. :1.0000
##	Snowparks Org	Snowparks	Nightskiing Org	Nightskiing
##	Length:480	Min. :0.0000	Length:480	Min. :0.0000
##	Class :character	1st Qu.:1.0000	Class :character	1st Qu.:0.0000
##	Mode :character	Median :1.0000	Mode :character	Median :0.0000
##		Mean :0.7521		Mean :0.3979
##		3rd Qu.:1.0000		3rd Qu.:1.0000
##		Max. :1.0000		Max. :1.0000
##	Summer skiing Org	Summer skiing		
##	Length:480	Min. :0.00000		
##	Class :character	1st Qu.:0.00000		
##	Mode :character	Median :0.00000		

```
##           Mean    :0.05833
##           3rd Qu.:0.00000
##           Max.    :1.00000
```

There is a file `mg_snow.csv` that contains data regarding snow cover for selected GPS locations for the entire 2022 year.

Name	Description	Notes
Month	Date to represent the month of the year (not just the first day)	
Latitude	Latitude at the center of the region (every “region” is 0.25x0.25 degrees in size)	
Longitude	Longitude at the center of the region (every “region” is 0.25x0.25 degrees in size)	
Snow	Percent of time the region was covered in snow during the month	

Processed file input/mg_snow.csv summary

```
##           Month           Latitude           Longitude           Snow
## Min.      :2022-01-01   Min.      : -55.38   Min.      : -179.88   Min.      :  0.39
## 1st Qu.   :2022-02-01   1st Qu.   : 50.88   1st Qu.   : -78.62   1st Qu.   : 59.45
## Median    :2022-04-01   Median    : 60.62   Median    :  36.88   Median    :100.00
## Mean      :2022-05-23   Mean      : 57.62   Mean      :  13.29   Mean      : 78.02
## 3rd Qu.   :2022-10-01   3rd Qu.   : 67.38   3rd Qu.   :  98.38   3rd Qu.   :100.00
## Max.      :2022-12-01   Max.      : 83.38   Max.      : 179.88   Max.      :100.00
```

4 Data Processing

4.1 Data Cleaning

I used mainly Google Sheets for data cleaning. The number of records is 480 and it was not an issue to process it. I used among others VLOOKUP, Data Cleanup

4.1.1 Duplicates

- The below records have all fields the same values, including latitude and longitude. I removed 143, 145, and 146, and renamed the resort field in ID=140 to “4 Vallees (Verbier, Nendaz, Thyon, Veysonnaz)”.

ID	Resort
140	Verbier (4 Vallees)
143	Nendaz (4 Vallees)
145	Thyon (4 Vallees)
146	Veysonnaz (4 Vallees)

- Records: ID=15 and ID=134 have the same values, including country and longitude and latitude. Interestingly, although Cervinia is in Italy the country record shows Switzerland so I assumed the record ID=15 should remain

15 Zermatt - Matterhorn 134 Cervinia

- Records: ID=104 and 111, have the same field values apart from ID, Resort, and Longest Run (Anton - 10, Lech - 6, which roughly is what I can find on-line: 11km and 4.5km, respectively. However, the total number of lifts for Lech is 34, for Anton is 38, and for Warth-Schroecken is 15, which gives total 87 as

of Mar 17, 2023. So I assume the data regards a single place Ski Arlberg that includes (Anton, Lech, and Warth-Schroecken). I remove ID=11, and rename St. Anton to Ski Arlberg (St Anton and Lech)

Also combining 120 and 123 records (Zurs, and St. Christoph)

ID	Resort	Longest Run
104	St. Anton (Ski Arlberg)	10
111	Lech (Ski Arlberg)	6

- Records: ID=20 and 32 have many the same field values, the name is just reverted. The resort is in France (not in Italy as in record ID=32) and the longest run in record ID=32 is 0. I removed ID=32.
- Records: ID 19, 61, 62 seems to have many same values, so I leave only one
- Records: ID 66, 67, 69, 74, 149, 150 have very same values, only only 66 has the longest run > 0 , so I leave it:
Les 3 Vallees: Courchevel, Saint Martin de Belleville, La Tania-Val Thorens/Les Menuires/Meribel, Val Thorens, Meribel, Les Menuires
- Records: ID 38, 72, and 133, seems values are same so inferring that the records have combined stats. Rec. 72 longest run has 0, so removing it.
- Records: 106 and 148 combined into 106 (148 removed)
- 118, 121 combined - record 121 removed.

4.1.2 Resort

I used GoogleSheets to deal with special characters. It seems that it is in Windows-1252 or Latin-1 encoding. I removed question marks.

4.1.3 Country

- 38 unique values

Country	No. of Resorts
Austria	84
USA	78
France	73
Switzerland	55
Italy	42

4.1.4 Continent

- 5 unique values

Continent	No. of Resorts
Europe	341
North America	98
Asia	24
Oceania	10
South America	7

4.1.5 Price

The below records have price = 0.

ID	Resort	Continent
196	Perisher	Oceania
235	Yellowstone Club	North America
335	Pragelato	Europe
349	Uludag-Bursa	Asia
369	Palandoken-Ejder 3200 World Ski Center-	Asia
387	High1 Resort	Asia
390	Alpika Service	Europe
475	Sun Mountain-Yabuli	Asia
495	Puigmal	Europe

4.1.6 Season

The season field can have a single range, two ranges, three ranges, unknown or year-around value. A single range can be a single month like “April” or “December-April”. Two and three month ranges does not include a single month. There are two three seasons value: “December - April, June - August, October - November”, and “October - November, December - May, June - October.” The second one accounts for the entire year. I transformed it with VLOOKUP and marked each month whether it is in season. If there are no gaps, i.e., no bordering ‘0’ then the information about when the season starts and ends is lost (i.e., where there is continuity of ‘1’). I can add a field a number of seasons, though.

4.1.7 Latitude and Longitude

The min latitude in $[0; 32.7558392] = 32.7558392$ and the max latitude in $[-30.0237148, 0) = -30.0237148$.

4.2 Missing Values

Surprisingly there were no missing values.

4.3 White Spaces

I checked with Google Sheets there are no white spaces.

5 Data Credibility

The data is provided by a respected website devoted to Data Analytics [1].

6 Data Bias

My assumption is that the data is collected and presented in a fair manner.

7 License

According to [1] the license is *Public Domain*.

8 Future Work

I have not used the data related to the snow cover. Including viz and info regarding the snow cover for resorts of interest could be valuable for ski trip planners.

References

- [1] M. Analytics, “Maven Challenge Ski Resorts.” Maven Analytics, Mar. 2023. Available: <https://app.mavenanalytics.io/datasets>
- [2] S. resorts stats & NASA Earth Observations, “Ski-resort-stats.com & NASA Earth Observations.” Mar. 2023. Available: <https://ski-resort-stats.com>