The excel file shows one way of analyzing the data. The goal is to significantly improve the methodology used in the excel file (improve the R-square calculated in the tab ‘REIT-Qtr’). Some variables you can play with include

* **Year over year cap/floor** - the raw data beneath the year over year change is apples to oranges… it includes the lowest available price at a hotel, which could be the price of a single room or a double room or a suite. In order to minimize the impact of the apples to oranges comparison, the year over change is capped.
* Using the year over year change at the closest hotels to improve the estimate for year over year change at a hotel under consideration (use the lat/long data).
* Use different weights for weekdays and weekends when calculating the average for the month or the quarter (excel file does a simple average)

Feel free to think more creatively on how else the r-square could be improved.

**Excel file:**

|  |  |
| --- | --- |
| **Tab** | **Description** |
|  |  |
| Analysis | Contains Year over Year change for each hotel\_id (column C) by date; column D includes Portfolio name that contains the hotel with id listed in column c; Column H has the weight of this particular hotel within the portfolio; **the Year over Year change is capped at ±20%** |
| AnalysisW | If the YearOverYear figure is a value, use the weight |
| AnalysisWAdj | Adjust the weights for each portfolio so that the sum of all weights is equal to 100% (essentially if some hotel does not have a year over year value on a particular date, its weight is redistributed to other hotels that have a year over year value within that portfolio. |
| AnalysisMkt-Count | Counts the number of hotels within a portfolio that have a year over year value on a specific date; Column D contains the number of hotels within the portfolio; Column C contains the minimum number of hotels that have to have a year over year value for us to use the portfolio average for that day for our analysis (if only 1 of 100 hotels have a year over year value on a certain day, then the change for that one hotel should not be used to calculate the overall portfolio average) |
| AnalysisMktD | If the count in ‘AnalysisMkt-Count’ is greater than the minimum count required for that portfolio, then calculate the weighted average change for the portfolio for that day – multiply the adjusted weight (tab - AnalysisWAdj) by year over year change (Analysis) |
| AnalysisMktM | Averages a portfolio change over a month – simple average over AnalysisMktD |
| AnalysisMktQ | Averages a portfolio change over a quarter – simple average over AnalysisMktD |
| REIT-Qtr | Calculates the r-square between the actual year over year change for the portfolio and what the excel file’s data analysis says (6 data-points) |
|  |  |

**Hotel\_trip\_reits** - includes information on portfolios (what hotels are included in these portfolios and what is the weight of each hotel in the portfolio)

**Hotel\_trip\_prices\_yy\_min** - database of year over year changes for a large selection of hotels (includes hotels that are not part of the portfolios included in hotel\_trip\_reits; info on these extra hotels is included so that info on closest hotels - using latitude/longitude coordinates - can be used to estimate the price change at the hotel under consideration.

**Hotel\_trip\_latlong** - includes latitude longitude coordinates of a large inventory of hotel (this could be used to calculate the distance between hotels)

* All hotels included in hotel\_trip\_latlong will not have year over year data in hotel\_trip\_prices\_yy\_min. Carefully select the list of closest hotels that also have considerable number of datapoints for year over year data.
* You could create the following function in mySQL / workbench and use it to calculate distance between two lat/long coordinates…

CREATE DEFINER=`root`@`localhost` FUNCTION `dist\_between`(lat1 DOUBLE, lon1 DOUBLE, lat2 DOUBLE, lon2 DOUBLE) RETURNS double

DETERMINISTIC

RETURN ACOS( SIN(lat1\*PI()/180)\*SIN(lat2\*PI()/180) + COS(lat1\*PI()/180)\*COS(lat2\*PI()/180)\*COS(lon2\*PI()/180-lon1\*PI()/180) ) \* 3949.9

**Hotel\_trip\_prices\_yy\_min.sql**

|  |  |
| --- | --- |
| hotel\_id | identifies a unique hotel (matches across tables) |
| yy\_change | Year over Year change in price  Price on 3/28/17 = 110  Price on 3/28/16 = 100  yy\_change = 110/100-1 = 10%  myDate = 2017-03-28 |

**hotel\_trip\_reits.sql**

|  |  |
| --- | --- |
| hotel\_id | identifies a unique hotel (matches across tables) |
| Ticker | Identifies a portfolio for which the year over year changes are aggregated. |
| Weights | Weight for the hotel\_id within a Portfolio (Ticker). These may or may not add up to 100%. In case, they do not, they weights are re-calculated (adjusted weights as shown in the excel file). |
| Latitude | Latitude of the hotel represented by hotel\_id |
| Latitude | Longitude of the hotel represented by hotel\_id |

**hotel\_trip\_latlong.sql**

|  |  |
| --- | --- |
| hotel\_id | identifies a unique hotel (matches across tables) |
| Latitude | Latitude of the hotel represented by hotel\_id |
| Latitude | Longitude of the hotel represented by hotel\_id |