**Пример данных:**

AVwc252WIN2L1WUfpqLP *(id)* – очевидно

2016-10-30T21:42:42Z (dateAdded) - Дата первого добавления этой компании в бизнес-базу данных.

2018-09-10T21:06:27Z *(dateUpdated)* - Самая последняя дата, когда эта компания была обновлена ​​или видна нашей системой.

5921 Valencia Cir *(address)* - Физический почтовый адрес этого офиса.

""Hotels,Hotels and motels,Hotel and motel reservations,Resorts,Resort,Hotel" *(categories)* - Список ключевых слов категории, используемых для этого бизнеса, из нескольких источников.

Accommodation & Food Services *(primaryCategories)* - Список стандартизированных категорий, к которым принадлежит этот бизнес. (Главная категория)

Rancho Santa Fe *(city)* - Город, в котором находится предприятие.

US *(country)* - Двухбуквенный код страны местонахождения компании.

us/ca/ranchosantafe/5921valenciacir/359754519 *(keys)* - Список внутренних идентификаторов Datafiniti для этой компании. keysПоле используется для объединения необработанных данных от отдельных источников в мастер Datafiniti записи.

32.990959 *(latitude)* - Координаты широты для местоположения компании.

-117.186136 *(longitude)* - Координаты долготы для этого места нахождения компании.

Rancho Valencia Resort Spa *(name)* - Название местонахождения компании.

92067 *(postalCode)* - Почтовый индекс местонахождения компании.

CA *(province)* - Провинция или штат, в котором находится предприятие.

2013-11-14T00:00:00Z *(*[*reviews.date*](https://vk.com/away.php?to=http%3A%2F%2Freviews.date&cc_key=)*)* - дата отзыва, видимо

"2016-08-03T00:00:00Z,2016-07-26T00:00:00Z,2016-08-02T00:00:00Z,2016-07-17T00:00:00Z,2016-08-27T00:00:00Z,2016-08-23T00:00:00Z,2016-08-09T00:00:00Z,2016-11-07T00:00:00Z,2016-08-26T00:00:00Z" *(reviews.dateSeen)* - даты просмотров, видимо

5 *(reviews.rating)* - оценка отеля из отзыва, видимо

[https://www.hotels.com/hotel/125419/reviews/](https://vk.com/away.php?to=https%3A%2F%2Fwww.hotels.com%2Fhotel%2F125419%2Freviews%2520%2F&cc_key=) *(reviews.sourceURLs)* - Список URL-адресов, используемых для создания данных об этом местонахождении компании.

Our experience at Rancho Valencia was absolutely perfect from beginning to end!!!! We felt special and very happy during our stayed. I would come back in a heart beat!!! *(reviews.text)* - текст отзыва, видимо

Best romantic vacation ever!!!! *(reviews.title)* - заголовок отзыва, видимо

,,, (???) *(reviews.userCity, reviews.userProvince)* - пропущенные значения

Paula *(reviews.username)* - имя пользователя

"http://[www.hotels.com/ho125419/%252525253Flocale%252525253De..](https://vk.com/away.php?to=http%3A%2F%2Fwww.hotels.com%2Fho125419%2F%2525252525253Flocale%2525252525253Den_US%25252525252526pos%2525252525253DHCOM_US&cc_key=),  
[http://www.yellowbook.com/profile/rancho-valencia-res..](https://vk.com/away.php?to=http%3A%2F%2Fwww.yellowbook.com%2Fprofile%2Francho-valencia-resort-and-spa_1895174051.html%2Chttps%3A%2F%2Fwww.hotels.com%2Fhotel%2F125419%2Freviews%2520%2F%2Chttps%3A%2F%2Ffoursquare.com%2Fv%2Francho-valencia-resort--spa%2F4aaedb79f964a5208f6320e3%2Chttp%3A%2F%2Fwww.gayot.com%2FHotels%2FSelect-Your-City%2FUnited-States%2FCalifornia%2FSan-Diego%2FRancho-Valencia-Resort-Spa-Rancho-Santa-Fe-SDHOT02319-02&cc_key=)" *(sourceURLs)* - Список URL-адресов, используемых для создания данных об этом местонахождении компании (мб список ссылок в отзыве)  
[http://www.ranchovalencia.com](https://vk.com/away.php?to=http%3A%2F%2Fwww.ranchovalencia.com&cc_key=) *(websites)* - Список сайтов для бизнеса.

**P.s.**  
[https://developer.datafiniti.co/docs/business-data-sc..](https://vk.com/away.php?to=https%3A%2F%2Fdeveloper.datafiniti.co%2Fdocs%2Fbusiness-data-schema&cc_key=) - для расшифровки

Предобработка:

* Невооруженным взглядом (первым тыком) было выявлено, что некоторая информация о пользователе умалчивается. Можно попробовать её поискать с сайтов (сложно, но плюсик), либо почистить. Информация типа имени пользователя и его местоположения

Что можно наанализировать с данными:

<------------------------------------------------------------------------------------------------------------------>

* Посмотреть на даты первого-последнего обновлений, можно сгруппировать и в целом посмотреть, какая попалась выборка: насколько актуальная информация (последние обновления) и насколько большой интервал времени охватывает (первое появление)
* Сгруппировать по физическим почтовым адресам (без цифорок) и построить гистограммку)
* Погруппировать по главным категориям и тоже вывести это в график (увидеть, на что у нас имеются отзывы. Можно при желании посмотреть количество отзывов на конкретную отрасль)
* Сгруппировать по городам

<------------------------------------------------------------------------------------------------------------------>

* Группировка по странам и вывод пары статистик по этому делу: число предприятий, суммарное число отзывов, средние оценки и тд
* Провинции, штаты - тоже группируем
* Отзывы: тоже группируем по областям, сайтам, откуда пришли и так далее. Вывести какие-нибудь статистики типа средней оценки.
* Можно по широте и долготе узнать,где находится больше всего отелей. Cкорее всего отелей больше к морю, в южной части мб побольше чем в северной
* Найти топ-10 городов, где больше всего высокооцененных отелей.
* <------------------------------------------------------------------------------------------------------------------>

* Запарности: можно попробовать построить модель, которая будет предсказывать оценку отзыва по присутствующим словам :D