The following mysql queries represent a large fraction (maybe a fourth) of the mysql queries that I made during exploratory analysis. I mostly collected the queries that I thought that I would want to use again later.

```
select user_id,
avg(booking),
stddev(distance),
avg(max_kids),
avg(max_adults),
avg(room_size),
avg(num_rooms) from
(select
        user_id,
        max(is_booking) booking,
        avg(orig_destination_distance) distance,
        max(srch_children_cnt) max_kids,
        max(srch_adults_cnt) max_adults,
        avg(srch_children_cnt + srch_adults_cnt) room_size,
        avg(srch_rm_cnt) num_rooms
        from Expedia group by user_id, date(date_time) order by count(user_id) desc)t
group by user_id;

select user_id, count(is_booking), avg(is_booking), max(srch_children_cnt) max_kids,
max(srch_adults_cnt) max_adults, avg(srch_children_cnt + srch_adults_cnt) avg_room_size,
avg(srch_rm_cnt) from Expedia group by user_id having avg(srch_rm_cnt) > 1.8;

select * from Expedia where Expedia.srch_destination_id in (
        select srch_destination_id from Dest where srch_destination_name like "New York, New
York, United States of America" or srch_destination_name like "New York% New York, United
States of America" or srch_destination_name like "%New York, New York, United States of
America");

select user_id, max(is_booking), count(is_booking) from NY group by user_id, srch_ci;

create table Nash(date_time datetime, site_name varchar(56), user_location_country
varchar(56),user_location_region varchar(56), user_location_city varchar(56),
user_location_latitude double, user_location_longitude double, orig_destination_distance
double, user_id int, is_mobile int, is_package int, channel int, srch_ci date, srch_co date,
srch_adults_cnt int, srch_children_cnt int, srch_rm_cnt int, srch_destination_id int, hotel_country
varchar(56), is_booking int, hotel_id int, prop_is_branded int, prop_starrating int, distance_band
varchar(12), hist_price_band varchar(12), popularity_band varchar(12), cnt int);

load data local infile '/Users/z/Desktop/datafest/nash_0.csv' into table Nash fields terminated by
',';

select * from Expedia where Expedia.srch_destination_id in (
        select srch_destination_id from Dest where srch_destination_name like "Seattle,
Washington, United States of America" or srch_destination_name like "Seattle% Washington,
```

```
United States of America" or srch_destination_name like "%Seattle, Washington, United States
of America");

select user_id, max(is_booking), count(is_booking) from Nash group by user_id, srch_ci;

select * from Expedia where Expedia.srch_destination_id in (
        select srch_destination_id from Dest where srch_destination_name like "San Diego,
California, United States of America" or srch_destination_name like "San Diego% California,
United States of America" or srch_destination_name like "%San Diego, California, United States
of America");

select * from Expedia where Expedia.srch_destination_id in (
        select srch_destination_id from Dest where srch_destination_name like "Miami, Florida,
United States of America" or srch_destination_name like "Miami% Florida, United States of
America" or srch_destination_name like "%Miami, Florida, United States of America");

select user_id, max(is_booking), count(is_booking), user_location_country,
user_location_region, user_location_city, user_location_latitude, user_location_longitude,
max(orig_destination_distance), avg(srch_adults_cnt), avg(srch_children_cnt),
max(srch_adults_cnt + srch_children_cnt), avg(srch_rm_cnt), max(srch_rm_cnt),
srch_destination_id, hotel_country, avg(prop_is_branded), avg(prop_starrating), sum(cnt) from
Sea group by user_id, srch_ci, user_location_country, user_location_region, user_location_city,
user_location_latitude, user_location_longitude, srch_destination_id, hotel_country;

select user_id, max(is_booking), count(is_booking), user_location_country,
user_location_region, user_location_city, user_location_latitude, user_location_longitude,
max(orig_destination_distance), avg(srch_adults_cnt), avg(srch_children_cnt),
max(srch_adults_cnt + srch_children_cnt), avg(srch_rm_cnt), max(srch_rm_cnt),
srch_destination_id, hotel_country, avg(prop_is_branded), avg(prop_starrating), sum(cnt) from
MI group by user_id, srch_ci, user_location_country, user_location_region, user_location_city,
user_location_latitude, user_location_longitude, srch_destination_id, hotel_country;


select
        user_id,
        max(is_booking),
        count(is_booking),
        user_location_country,
        user_location_region,
        user_location_city,
        user_location_latitude,
        user_location_longitude,
        max(orig_destination_distance),
        avg(srch_adults_cnt),
        avg(srch_children_cnt),
        max(srch_adults_cnt + srch_children_cnt),
        avg(srch_rm_cnt),
        max(srch_rm_cnt),
        srch_destination_id,
```

```
        hotel_country,
        avg(prop_is_branded),
        avg(prop_starrating),
        sum(cnt)
from Nash
group by
        user_id,
        srch_ci,
        user_location_country,
        user_location_region,
        user_location_city,
        user_location_latitude,
        user_location_longitude,
        srch_destination_id,
        hotel_country;
```