Project 2 Part 2

Project:

a. Include name of the team members.

Alan Xu, Jai Bansal

b. Who is responsible for what? 1-2 sentences against each team member.

At this stage of the project, we are each responsible for working together on designing our schema, tables, and overall database and code layout.

Individually, we are both responsible for setting up our local development environments so that all of the data, tables, and code is the same (with version control via Git). The front-end setup will also be completed together (pair programming).

c. How are you sharing codebase? 1-2 sentences.

After setting up local development environments to be the same (same ports, package versions, IDE, configurations, etc.), we will use GitHub to share code between one another. For raw data, due to the size and perhaps sensitivity of the files, we will each download those locally instead of sharing on GitHub.

Database:

d. By this assignment submission date, you should have your raw data set imported in MySQL server in a megatable. You should also have started working on decomposition.

   Include screenshots of CREATE TABLE statements that you have completed/working on. This includes your megatable (required) and any other tables that you have designed.

The megatable for this project includes data for each pitch during every at bat, while also containing details about each at bat for every pitch. The obvious decomposition for us here was to decompose into a table that contains pitch data (pitches) and a table that contains at bat data (atbats). These two tables are joined on the foreign key ab_id, which is also the primary key for the atbats table. All fields in the atbats table are functionally dependent upon ab_id. The screenshots of how we created our megatable, as well as our decomposed tables are included below:

CREATE TABLE statement for `pitch_data` megatable (Screenshot below)

```
-- Create table `pitch_data`.`pitch_data`
DROP TABLE IF EXISTS `pitch_data`.`pitch_data` ;

CREATE TABLE IF NOT EXISTS `pitch_data`.`pitch_data` (
  `ab_id` INT(11) NOT NULL,
  `batter_id` INT(11) NULL,
  `event` VARCHAR(20) NULL,
  `g_id` INT(11) NULL,
  `inning` TINYINT NULL,
  `o` TINYINT NULL,
  `p_score` TINYINT NULL,
  `p_throws` CHAR(2) NULL,
  `pitcher_id` INT(11) NULL,
  `stand` CHAR(2) NULL,
  `top` VARCHAR(10) NULL,
  `px` DECIMAL(4,3) NULL,
  `pz` DECIMAL(4,3) NULL,
  `start_speed` DECIMAL(10,2) NULL,
  `end_speed` DECIMAL(10,2) NULL,
  `spin_rate` DECIMAL(10,3) NULL,
  `spin_dir` DECIMAL(10,3) NULL,
  `break_angle` DECIMAL(3,1) NULL,
  `break_length` DECIMAL(3,1) NULL,
  `break_y` DECIMAL(3,1) NULL,
  `ax` DECIMAL(6,3) NULL,
  `ay` DECIMAL(6,3) NULL,
  `az` DECIMAL(6,3) NULL,
  `sz_bot` DECIMAL(4,2) NULL,
  `sz_top` DECIMAL(4,2) NULL,
  `type_confidence` DECIMAL(4,3) NULL,
  `vx0` DECIMAL(6,3) NULL,
  `vy0` DECIMAL(6,3) NULL,
  `vz0` DECIMAL(6,3) NULL,
  `x` DECIMAL(10,3) NULL,
  `x0` DECIMAL(10,3) NULL,
  `y` DECIMAL(10,3) NULL,
  `y0` DECIMAL(10,3) NULL,
  `z0` DECIMAL(10,3) NULL,
  `pfx_x` DECIMAL(6,2) NULL,
  `pfx_z` DECIMAL(6,2) NULL,
  `nasty` INT(5) NULL,
  `zone` INT(5) NULL,
  `code` CHAR(3) NULL,
  `type` CHAR(2) NULL,
  `pitch_type` CHAR(3) NULL,
  `event_num` INT(5) NULL,
  `b_score` TINYINT NULL,
  `b_count` TINYINT NULL,
  `s_count` TINYINT NULL,
  `outs` TINYINT NULL,
  `pitch_num` TINYINT NULL,
  `on_1b` TINYINT NULL,
  `on_2b` TINYINT NULL,
  `on_3b` TINYINT NULL)
ENGINE = InnoDB;
```

CREATE TABLE statement for `pitches` (Screenshots below)

```
-- Create table `pitch_data`.`pitches`
DROP TABLE IF EXISTS `pitch_data`.`pitches` ;
CREATE TABLE IF NOT EXISTS `pitch_data`.`pitches` (
  `px` DECIMAL(4,3) NULL,
  `pz` DECIMAL(4,3) NULL,
  `start_speed` DECIMAL(10,2) NULL,
  `end_speed` DECIMAL(10,2) NULL,
  `spin_rate` DECIMAL(10,3) NULL,
  `spin_dir` DECIMAL(10,3) NULL,
  `break_angle` DECIMAL(3,1) NULL,
  `break_length` DECIMAL(3,1) NULL,
  `break_y` DECIMAL(3,1) NULL,
  `ax` DECIMAL(6,3) NULL,
  `ay` DECIMAL(6,3) NULL,
  `az` DECIMAL(6,3) NULL,
  `sz_bot` DECIMAL(6,2) NULL,
  `sz_top` DECIMAL(6,2) NULL,
  `type_confidence` DECIMAL(4,3) NULL,
  `vx0` DECIMAL(6,3) NULL,
  `vy0` DECIMAL(6,3) NULL,
  `vz0` DECIMAL(6,3) NULL,
  `x` DECIMAL(10,3) NULL,
  `x0` DECIMAL(10,3) NULL,
  `y` DECIMAL(10,3) NULL,
  `y0` DECIMAL(10,3) NULL,
  `z0` DECIMAL(10,3) NULL,
  `pfx_x` DECIMAL(6,2) NULL,
  `pfx_z` DECIMAL(6,2) NULL,
  `nasty` INT(5) NULL,
  `zone` INT(5) NULL,
  `code` CHAR(3) NULL,
  `type` CHAR(2) NULL,
  `pitch_type` CHAR(3) NULL,
  `event_num` INT(5) NULL,
  `b_score` TINYINT NULL,
  `ab_id` INT(11) NULL,
  `b_count` TINYINT NULL,
  `s_count` TINYINT NULL,
  `outs` TINYINT NULL,
  `pitch_num` TINYINT NULL,
  `on_1b` TINYINT NULL,
  `on_2b` TINYINT NULL,
  `on_3b` TINYINT NULL,
  INDEX `fk_ab_id_idx` (`ab_id` ASC),
  CONSTRAINT `fk_ab_id`
    FOREIGN KEY (`ab_id`)
    REFERENCES `pitch_data`.`atbats` (`ab_id`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB;
```

CREATE TABLE statement for `atbats` (Screenshot below)

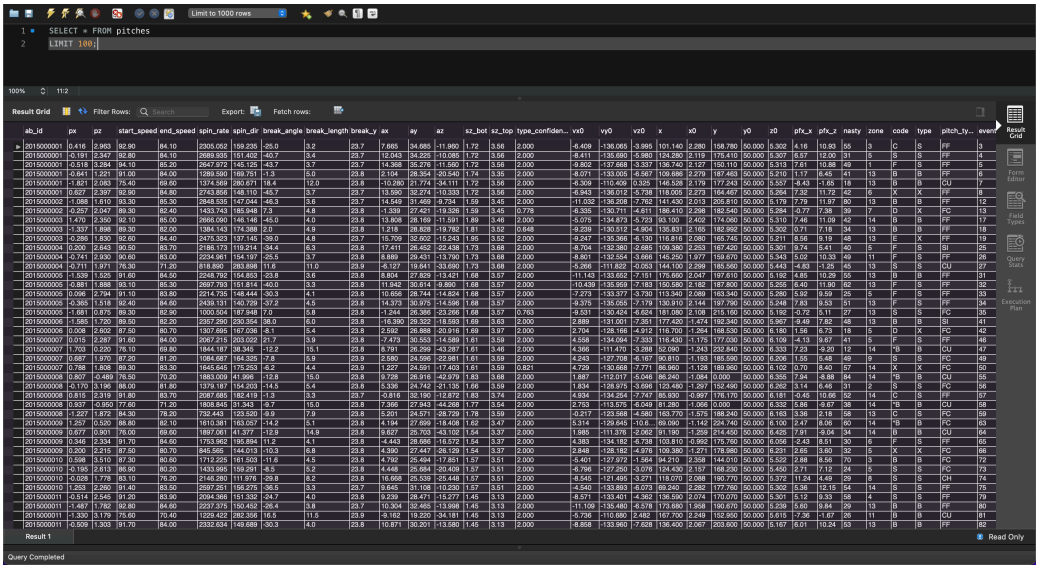
```
DROP TABLE IF EXISTS `pitch_data`.`atbats` ;

CREATE TABLE IF NOT EXISTS `pitch_data`.`atbats` (
  `ab_id` INT(11) NOT NULL,
  `batter_id` INT(11) NULL,
  `event` VARCHAR(20) NULL,
  `g_id` INT(11) NULL,
  `inning` TINYINT NULL,
  `o` TINYINT NULL,
  `p_score` TINYINT NULL,
  `p_throws` CHAR(2) NULL,
  `pitcher_id` INT(11) NULL,
  `stand` CHAR(2) NULL,
  `top` VARCHAR(10) NULL,
  PRIMARY KEY (`ab_id`))
ENGINE = InnoDB;
```

Include screenshot of fetched data from your megatable by running a simple query like following (required):

SELECT * FROM megatable LIMIT 100;

Query of megatable



Include screenshot of any other relevant code blocks. Clearly label the screenshots.

LOAD DATA statements to populate the database from .csv files

```
USE `pitch_data`;

-- Load atbat data from atbat.csv
LOAD DATA
    INFILE '/Users/alanxu/mlb_pitch_data/raw_data/atbats.csv'
    INTO TABLE atbats
    FIELDS TERMINATED BY ','
    ENCLOSED BY '"'
    LINES TERMINATED BY '\n'
    IGNORE 1 ROWS; -- Header Line


-- Load pitch data from pitches.csv
LOAD DATA
    INFILE '/Users/alanxu/mlb_pitch_data/raw_data/pitches.csv'
    INTO TABLE pitches
    FIELDS TERMINATED BY ','
    ENCLOSED BY '"'
    LINES TERMINATED BY '\n'
    IGNORE 1 ROWS; -- Header line
```

e. In terms of percentage how much you think you have completed on database side of the project? Describe the completed work in 2-3 sentences. Describe roadblocks, if any.

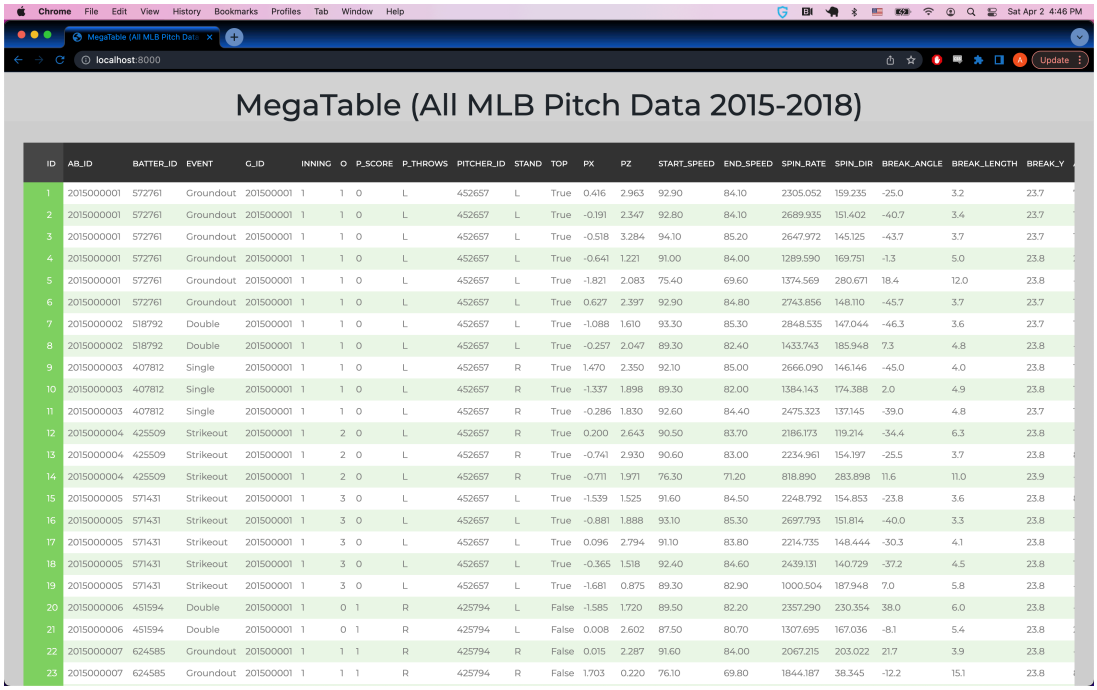
Since we did the database modeling portion of the assignment together, I would say I contributed 50% to the database side of the project. This involved designing the schemas, tables (with appropriate data types), writing scripts for populating the tables, decomposing, and verifying that the data is valid. The biggest roadblocks were definitely dealing with inconsistencies in the raw data, which need to be handled during table insertion.

Front end:

f. By this assignment submission date, you should have decided on the front-end application programming language and have a successful connection established between the front end and the project database.

Attach a screenshot of the browser showing a successful connection displaying some data from any table of your project (can be megatable). Clearly label the screenshots.

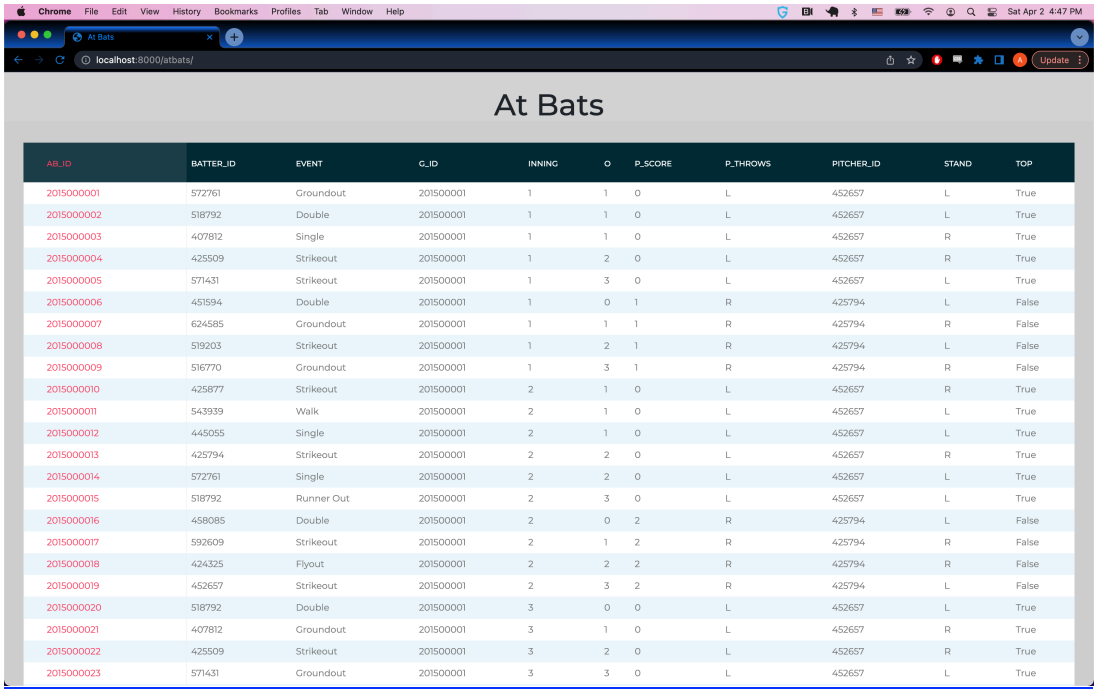
Screenshot of Megatable data displayed on the front-end client (browser):



MegaTable (All MLB Pitch Data 2015-2018)

ID	AB_ID	BATTER_ID	EVENT	G_ID	INNING	O	P_SCORE	P_THROWS	PITCHER_ID	STAND	TOP	PX	PZ	START_SPEED	END_SPEED	SPIN_RATE	SPIN_DIR	BREAK_ANGLE	BREAK_LENGTH	BREAK_Y
1	2015000001	572761	Groundout	201500001	1	1	0	L	452657	L	True	0.416	2.963	92.90	84.10	2305.052	159.235	-25.0	3.2	23.7
2	2015000001	572761	Groundout	201500001	1	1	0	L	452657	L	True	-0.191	2.347	92.80	84.10	2689.935	151.402	-40.7	3.4	23.7
3	2015000001	572761	Groundout	201500001	1	1	0	L	452657	L	True	-0.518	3.284	94.10	85.20	2647.972	145.125	-43.7	3.7	23.7
4	2015000001	572761	Groundout	201500001	1	1	0	L	452657	L	True	-0.641	1.221	91.00	84.00	1289.590	169.751	-1.3	5.0	23.8
5	2015000001	572761	Groundout	201500001	1	1	0	L	452657	L	True	-1.821	2.083	75.40	69.60	1374.569	280.671	18.4	12.0	23.8
6	2015000001	572761	Groundout	201500001	1	1	0	L	452657	L	True	0.627	2.397	92.90	84.80	2743.856	148.110	-45.7	3.7	23.7
7	2015000002	518792	Double	201500001	1	1	0	L	452657	L	True	-1.088	1.610	93.30	85.30	2848.535	147.044	-46.3	3.6	23.7
8	2015000002	518792	Double	201500001	1	1	0	L	452657	L	True	-0.257	2.047	89.30	82.40	1433.743	185.948	7.3	4.8	23.8
9	2015000003	407812	Single	201500001	1	1	0	L	452657	R	True	1.470	2.350	92.10	85.00	2666.090	146.146	-45.0	4.0	23.8
10	2015000003	407812	Single	201500001	1	1	0	L	452657	R	True	-1.337	1.898	89.30	82.00	1384.143	174.388	2.0	4.9	23.8
11	2015000003	407812	Single	201500001	1	1	0	L	452657	R	True	-0.286	1.830	92.60	84.40	2475.323	137.145	-39.0	4.8	23.7
12	2015000004	425509	Strikeout	201500001	1	2	0	L	452657	R	True	0.200	2.643	90.50	83.70	2186.173	119.214	-34.4	6.3	23.8
13	2015000004	425509	Strikeout	201500001	1	2	0	L	452657	R	True	-0.741	2.930	90.60	83.00	2234.961	154.197	-25.5	3.7	23.8
14	2015000004	425509	Strikeout	201500001	1	2	0	L	452657	R	True	-0.711	1.971	76.30	71.20	818.890	283.898	11.6	11.0	23.9
15	2015000005	571431	Strikeout	201500001	1	3	0	L	452657	L	True	-1.539	1.525	91.60	84.50	2248.792	154.853	-23.8	3.6	23.8
16	2015000005	571431	Strikeout	201500001	1	3	0	L	452657	L	True	-0.881	1.888	93.10	85.30	2697.793	151.814	-40.0	3.3	23.8
17	2015000005	571431	Strikeout	201500001	1	3	0	L	452657	L	True	0.096	2.794	91.10	83.80	2214.735	148.444	-30.3	4.1	23.8
18	2015000005	571431	Strikeout	201500001	1	3	0	L	452657	L	True	-0.365	1.518	92.40	84.60	2439.131	140.729	-37.2	4.5	23.8
19	2015000005	571431	Strikeout	201500001	1	3	0	L	452657	L	True	-1.681	0.875	89.30	82.90	1000.504	187.948	7.0	5.8	23.8
20	2015000006	451594	Double	201500001	1	0	1	R	425794	L	False	-1.585	1.720	89.50	82.20	2357.290	230.354	38.0	6.0	23.8
21	2015000006	451594	Double	201500001	1	0	1	R	425794	L	False	0.008	2.602	87.50	80.70	1307.695	167.036	-8.1	5.4	23.8
22	2015000007	624585	Groundout	201500001	1	1	1	R	425794	R	False	0.015	2.287	91.60	84.00	2067.215	203.022	21.7	3.9	23.8
23	2015000007	624585	Groundout	201500001	1	1	1	R	425794	R	False	1.703	0.220	76.10	69.80	1844.187	38.345	-12.2	15.1	23.8

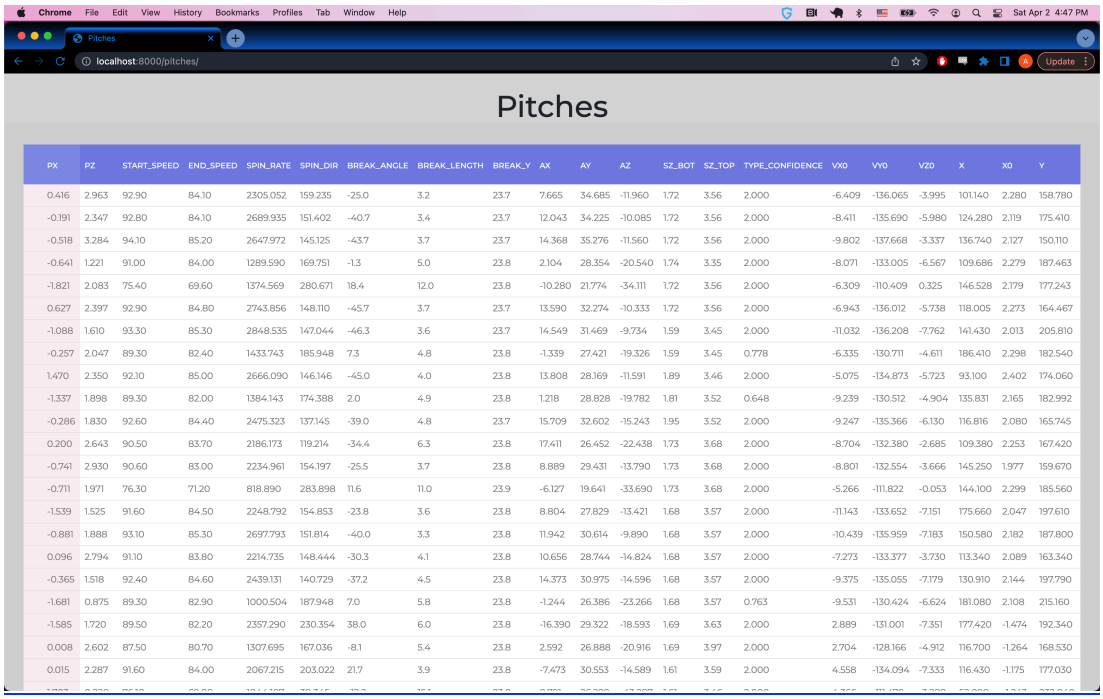
Screenshot of At Bat data displayed on the front-end client (browser):



At Bats

AB_ID	BATTER_ID	EVENT	G_ID	INNING	O	P_SCORE	P_THROWS	PITCHER_ID	STAND	TOP
2015000001	572761	Groundout	201500001	1	1	0	L	452657	L	True
2015000002	518792	Double	201500001	1	1	0	L	452657	L	True
2015000003	407812	Single	201500001	1	1	0	L	452657	R	True
2015000004	425509	Strikeout	201500001	1	2	0	L	452657	R	True
2015000005	571431	Strikeout	201500001	1	3	0	L	452657	L	True
2015000006	451594	Double	201500001	1	0	1	R	425794	L	False
2015000007	624585	Groundout	201500001	1	1	1	R	425794	R	False
2015000008	519203	Strikeout	201500001	1	2	1	R	425794	L	False
2015000009	516770	Groundout	201500001	1	3	1	R	425794	R	False
2015000010	425877	Strikeout	201500001	2	1	0	L	452657	R	True
2015000011	543939	Walk	201500001	2	1	0	L	452657	L	True
2015000012	445055	Single	201500001	2	1	0	L	452657	L	True
2015000013	425794	Strikeout	201500001	2	2	0	L	452657	R	True
2015000014	572761	Single	201500001	2	2	0	L	452657	L	True
2015000015	518792	Runner Out	201500001	2	3	0	L	452657	L	True
2015000016	458085	Double	201500001	2	0	2	R	425794	L	False
2015000017	592609	Strikeout	201500001	2	1	2	R	425794	R	False
2015000018	424325	Flyout	201500001	2	2	2	R	425794	R	False
2015000019	452657	Strikeout	201500001	2	3	2	R	425794	L	False
2015000020	518792	Double	201500001	3	0	0	L	452657	L	True
2015000021	407812	Groundout	201500001	3	1	0	L	452657	R	True
2015000022	425509	Strikeout	201500001	3	2	0	L	452657	R	True
2015000023	571431	Groundout	201500001	3	3	0	L	452657	L	True

Screenshot of Pitches data displayed on the front-end client (browser):



PX	PZ	START_SPEED	END_SPEED	SPIN_RATE	SPIN_DIR	BREAK_ANGLE	BREAK_LENGTH	BREAK_Y	AX	AY	AZ	SZ_BOT	SZ_TOP	TYPE_CONFIDENCE	VX0	VY0	VZ0	X	X0	Y
0.416	2.963	92.90	84.10	2305.052	159.235	-25.0	3.2	23.7	7.665	34.685	-11.960	1.72	3.56	2.000	-6.409	-136.065	-3.995	101.140	2.280	158.780
-0.191	2.347	92.80	84.10	2689.935	151.402	-40.7	3.4	23.7	12.043	34.225	-10.085	1.72	3.56	2.000	-8.411	-135.690	-5.980	124.280	2.119	175.410
-0.518	3.284	94.10	85.20	2647.972	145.125	-43.7	3.7	23.7	14.368	35.276	-11.560	1.72	3.56	2.000	-9.802	-137.668	-3.337	136.740	2.127	150.110
-0.641	1.221	91.00	84.00	1289.590	169.751	-1.3	5.0	23.8	2.104	28.354	-20.540	1.74	3.35	2.000	-8.071	-133.005	-6.567	109.686	2.279	187.463
-1.821	2.083	75.40	69.60	1374.569	280.671	18.4	12.0	23.8	-10.280	21.774	-34.111	1.72	3.56	2.000	-6.309	-110.409	0.325	146.528	2.179	177.243
0.627	2.397	92.90	84.80	2743.856	148.110	-45.7	3.7	23.7	13.590	32.274	-10.333	1.72	3.56	2.000	-6.943	-136.012	-5.738	118.005	2.273	164.467
-1.088	1.610	93.30	85.30	2848.535	147.044	-46.3	3.6	23.7	14.549	31.469	-9.734	1.59	3.45	2.000	-11.032	-136.208	-7.762	141.430	2.013	205.810
-0.257	2.047	89.30	82.40	1433.743	185.948	7.3	4.8	23.8	-1.339	27.421	-19.326	1.59	3.45	0.778	-6.335	-130.711	-4.611	186.410	2.298	182.540
1.470	2.350	92.10	85.00	2666.090	146.146	-45.0	4.0	23.8	13.808	28.169	-11.591	1.89	3.46	2.000	-5.075	-134.873	-5.723	93.100	2.402	174.060
-1.337	1.898	89.30	82.00	1384.143	174.388	2.0	4.9	23.8	1.218	28.828	-19.782	1.81	3.52	0.648	-9.239	-130.512	-4.904	135.831	2.165	182.992
-0.286	1.830	92.60	84.40	2475.323	137.145	-39.0	4.8	23.7	15.709	32.602	-15.243	1.95	3.52	2.000	-9.247	-135.366	-6.130	116.816	2.080	165.745
0.200	2.643	90.50	83.70	2186.173	119.214	-34.4	6.3	23.8	17.411	26.452	-22.438	1.73	3.68	2.000	-8.704	-132.380	-2.685	109.380	2.253	167.420
-0.741	2.930	90.60	83.00	2234.961	154.197	-25.5	3.7	23.8	8.889	29.431	-13.790	1.73	3.68	2.000	-8.801	-132.554	-3.666	145.250	1.977	159.670
-0.711	1.971	76.30	71.20	818.890	283.898	11.6	11.0	23.9	-6.127	19.641	-33.690	1.73	3.68	2.000	-5.266	-111.822	-0.053	144.100	2.299	185.560
-1.539	1.525	91.60	84.50	2248.792	154.853	-23.8	3.6	23.8	8.804	27.829	-13.421	1.68	3.57	2.000	-11.143	-133.652	-7.151	175.660	2.047	197.610
-0.881	1.888	93.10	85.30	2697.793	151.814	-40.0	3.3	23.8	11.942	30.614	-9.890	1.68	3.57	2.000	-10.439	-135.959	-7.183	150.580	2.182	187.800
0.096	2.794	91.10	83.80	2214.735	148.444	-30.3	4.1	23.8	10.656	28.744	-14.824	1.68	3.57	2.000	-7.273	-133.377	-3.730	113.340	2.089	163.340
-0.365	1.518	92.40	84.60	2439.131	140.729	-37.2	4.5	23.8	14.373	30.975	-14.596	1.68	3.57	2.000	-9.375	-135.055	-7.179	130.910	2.144	197.790
-1.681	0.875	89.30	82.90	1000.504	187.948	7.0	5.8	23.8	-1.244	26.386	-23.266	1.68	3.57	0.763	-9.531	-130.424	-6.624	181.080	2.108	215.160
-1.585	1.720	89.50	82.20	2357.290	230.354	38.0	6.0	23.8	-16.390	29.322	-18.593	1.69	3.63	2.000	2.889	-131.001	-7.351	177.420	-1.474	192.340
0.008	2.602	87.50	80.70	1307.695	167.036	-8.1	5.4	23.8	2.592	26.888	-20.916	1.69	3.97	2.000	2.704	-128.166	-4.912	116.700	-1.264	168.530
0.015	2.287	91.60	84.00	2067.215	203.022	21.7	3.9	23.8	-7.473	30.553	-14.589	1.61	3.59	2.000	4.558	-134.094	-7.333	116.430	-1.175	177.030

g. Which front end application programming language are you working with? Has anyone from team has prior knowledge of working with front end? Has anyone from team has prior knowledge of working with the chosen front end application programming language. 2-3 sentences.

We are using Python with Django to build the front-end client. While both of us have used Python in the past, this is the first time we are using the Django framework as well as using MySQL with Python. Alan has front-end experience in the past, mostly pertaining to styling JavaScript web applications; therefore, working with Python will present an equal learning curve for both of us.

h. What is the status of front-end application? In terms of percentage how much you think you have completed on front-end side of the project?

Currently, the front-end application is able to retrieve data from the MySQL API in our app (which provides basic CRUD operations), and display them in tabular form on different views, which are served at their own routes (i.e. At Bat data is pulled from the atbats table, and rendered at <url>/atbats/<:page_number>). Each page renders 50 results for each table.

We also completed the front-end client together via pair programming, so we each contributed ~50% of the work for the initial client setup.

Next deliverable

- What are your next steps? By next deliverable (in 10 days) what do you plan to complete? Define clear goals.

The next big deliverable will be actually putting the data to use (i.e. analytics), while also adding additional code for additional functional requirements of the site (e.g. navigation bar, routability, etc.).

Submission:

Complete this document and save it as pdf. You must submit a PDF file named **p2-part2-lastname1-lastname2.pdf** (For example if I submit this document with John Smith, I would name it p2-part2-singh-smith.pdf). Submit your files on Brightspace.

You must include your name and your partner name in the Brightspace submission text box.

Each member of the team must make the submission of same file.

Grading:

This Assignment will be graded on the following criteria:

1. Completeness of document.
2. Completeness of required components at this stage of project.
3. Clear evidence of work completed.

NO grading will be done on file/s sent through email or not uploaded to Brightspace.