

Report

November 17, 2022

1 Typeracer dataset report

The dataset was datamined on 2022/11/16 from <https://play.typeracer.com/>

Typeracer is an online typing game, where multiple players are given a text and the task is to (correctly) type it as fast as possible. Overall goal of this game is to improve ones typing skills - speed and accuracy.

Data collected was limited to english texts, so that analysis can done on the QWERTY keyboard.

Each row of the dataset represents one match played by one player. The dataset consists of 8543 rows and 12 columns: - player: name of the player - match: n-th match the player has ever played - wpm: words per minute, a measure of typing speed - pos: at which position the player has placed (first=won, second, etc.) - acc: percentage of letters that were correctly written on the first try - letters: letters the text consisted of - latencies: time delay before each letter was correctly typed - mistakes: mistakes signifies whether letter was written correctly at the first try or not - letters_all: all letters typed and deleted (includes mistakes) - latencies_all: time delay before all keypresses - operations: operations[i] signifies whether letters_all[i] was: correctly typed, typed a mistake, deleted a mistake, deleted a correct letter - durations: how long each keypress lasted (keyboard buttons continuously send signal when pressed down)

(does not consider the existence of the shift key) shift-key information is unfortunately not included

Between 10%-20% of all replays were thrown out to keep the data clean - some parts of the typeracer replay have bugs (eg. characters get swapped, but the app somehow pretends it didn't happened) - some rarer parts of the replay encoding evade my understanding - even though the text may be in english, a few players use non-english layouts and can mistakenly write a non-ascii char; this would make it hard to say what key was pressed - detecting the keyboard layout is nearly imposible, unless the player specifies what layout they are using; in a case that the layout is not QWERTY (eg. AZERTY, dvorak, etc.) no replay is collected for that player

There are no missing values. Match text (the column 'letters') is duplicated for each player in the match.

1.1 Column types

player	string
match	int32
wpm	int32
acc	float64
pos	int32

```

letters          object
latencies        object
mistakes         object
letters_all      object
latencies_all    object
durations        object
operations       object
dtype: object

```

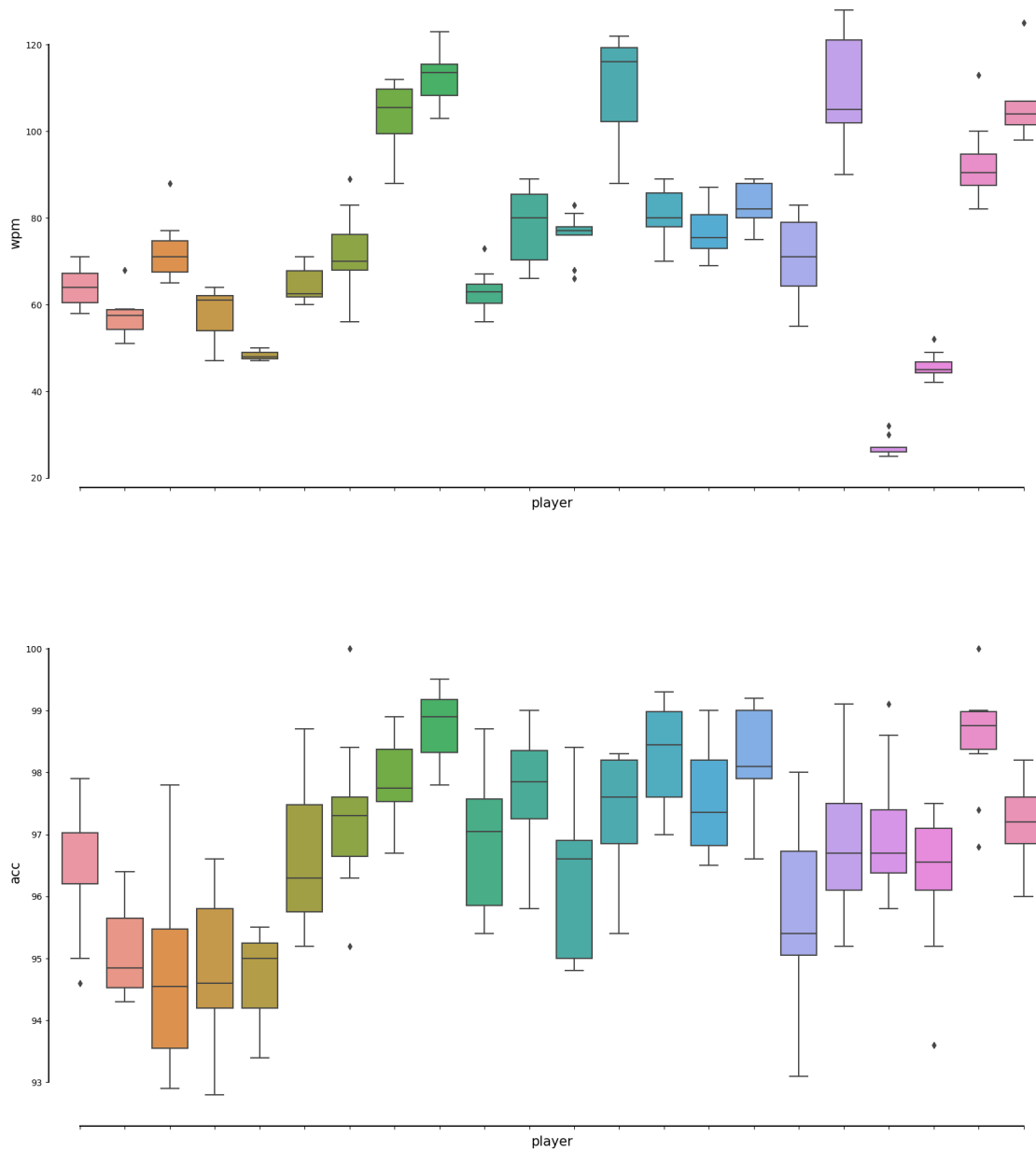
- letters, letters_all are ndarrays of char
- durations, latencies, latencies_all, durations are ndarrays of int
- mistakes is an ndarray of bool
- operations is an ndarray of values 0-3, categorical

	match	wpm	acc	pos
count	8543.000000	8543.000000	8543.000000	8543.000000
mean	15691.801007	79.625658	96.698069	2.364158
std	73769.434460	28.010164	1.827650	1.221256
min	1.000000	10.000000	85.700000	1.000000
25%	188.000000	61.000000	95.700000	1.000000
50%	936.000000	78.000000	96.900000	2.000000
75%	4537.500000	95.000000	98.000000	3.000000
max	630574.000000	244.000000	100.000000	7.000000

1.1.1 Hypotheses

- mistakes cost a lot of time since they require deleting and retyping => lots of keypresses; so there should be clear correlation between WPM and accuracy
- the further away a letter is from the neutral hand position the bigger the latency as there is more movement required
- more frequently used buttons probably have strongly developed muscle memory compared to seldomly used buttons
- letters that require using shift to type have larger latency (as total of 2 keypresses are required) and better accuracy (less speed related mistakes; oftentimes as the first letter at the start of a sentence)
- frequently switching hands while might have bad repercussions on accuracy due to synchronisation problems
- not switching hands frequently enough might lead to lower speed, because fingers can't 'prepare' to type the next letter and are required to perform lots of (possibly uncomfortable) motions without rest
- thumbs and pinkies are less nimble than other fingers, which probably leads to longer press duration
- in longer texts, the player gets progressively more and more tired, so it is expected that wpm and accuracy go down while latency goes up
- all those shortcomings might get lessened with experience
- high WPM players are more likely to win

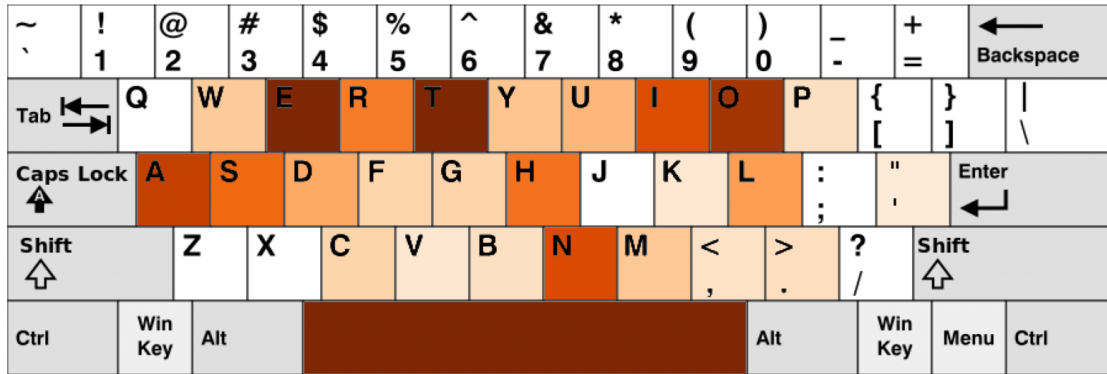
1.2 Player consistency



Some players seem to be more consistent than others. This applies to both wpm and accuracy.

1.3 Keyboard analytics

Rates of using a key:



This should basically mimic the distribution of letters in english.

Rates of mistyping a key:



This visualisation seems to be supporting the hypothesis, that less frequent letters are more likely to get mistyped

Mistake latencies:



Mistakes hurt similarly hard on all buttons. Probably because what takes time at a mistake is that the player continues typing for a while until realising he should stop. What then follows is a lot of deleting, which is also slow.

Average latencies:







~ `	1 !	2 @	3 #	4 \$	5 %	6 ^	7 &	8 *	9 (0)	- _	+ =	Backspace ←		
Tab ⇄	Q	W	E	R	T	Y	U	I	O	P	{ [}]	 \ _		
Caps Lock ⬆	A	S	D	F	G	H	J	K	L	;	: " ' ,	Enter ⬇			
Shift ⬆	Z	X	C	V	B	N	M	< ,	> .	? /	Shift ⬆				
Ctrl	Win Key	Alt										Alt	Win Key	Menu	Ctrl

Less frequent letters have higher latencies. Visualisation looks pretty much identical with the one with mistake rates.

Average lowercase latencies:

~ `	1 !	2 @	3 #	4 \$	5 %	6 ^	7 &	8 *	9 (0)	- _	+ =	Backspace ←	
Tab ↔	Q	W	E	R	T	Y	U	I	O	P	{ [}]	 \ ↵	
Caps Lock ⬆	A	S	D	F	G	H	J	K	L	⋮ ⋮ ⋮	" '	Enter ↵		
Shift ⬆	Z	X	C	V	B	N	M	< ,	> .	? /	Shift ⬆			
Ctrl	Win Key	Alt									Alt	Win Key	Menu	Ctrl

Average uppercase latencies:

~ 1	! 2	@ 3	# 4	\$ 5	% 6	^ 7	& 8	* 9	(0) -	+ =	 Backspace	
Tab 	Q	W	E	R	T	Y	U	I	O	P	{ [}]	 \ ~
Caps Lock 	A	S	D	F	G	H	J	K	L	: ;	" '	Enter 	
Shift 		Z	X	C	V	B	N	M	< ,	> .	? /	Shift 	
Ctrl	Win Key	Alt								Alt	Win Key	Menu	Ctrl

Latency difference between shifted and unshifted:

~ 1	! 2	@ 3	# 4	\$ 5	% 6	^ 7	& 8	* 9	(0) -	+ =	Backspace	
Tab ↔	Q	W	E	R	T	Y	U	I	O	P	{ [}]	 \ ~
Caps Lock ⬆	A	S	D	F	G	H	J	K	L	: ;	" '	Enter ⬇	
Shift ⬆	Z	X	C	V	B	N	M	< ,	> .	? /	Shift ⬆		
Ctrl	Win Key	Alt							Alt	Win Key	Menu	Ctrl	

1.4 WPM and accuracy correlation

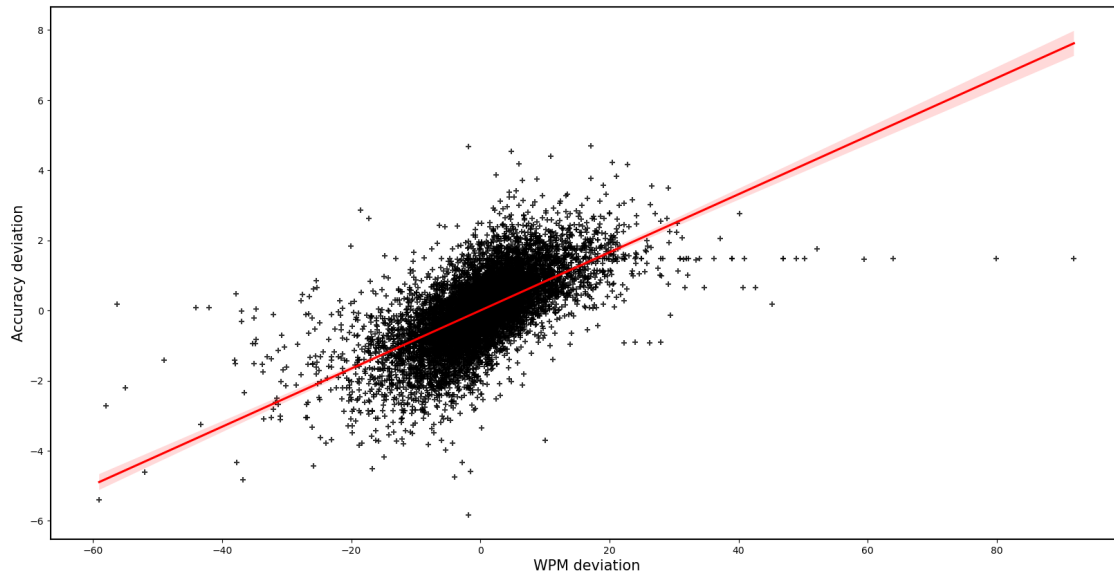
slope: 0.0828884382228181

intercept: 4.984835247978443e-16

r_value: 0.641695287487051

p_value: 0.0

std_err: 0.001071971605265329



Deviation of accuracy and deviation of wpm are strongly positively correlated

=> the less mistakes, the less time wasted repairing them

=> players should be mainly focusing on improving accuracy rather than raw speed