REPORT

BUSINESS INTELLIGENCE WORKPLACE

PROJECT P01: DESCRIPTIVE ANALYTICS

Author: Joanna Stawik

Group: Fridays 11:15

Table of Contents

1 INTRODUCTION

1.1 BACKGROUND OF THE PROJECT

The rapid urbanization has led to various life-threatening disasters caused by overcrowding. To effectively manage crowds, it is essential to understand the dynamic behavior of a crowd. Experimental data involving live human participants is necessary for crowd motion studies. In this project, the goal is to study real data of crossing flows of pedestrian crowds. The data was collected in Rennes, France in 2016 by a group of researchers from various fields, such as physics, computer science, and psychology.

During the experimental trials, participants were divided into two groups and asked to cross through each other to generate specific crossing angles. The trials were performed for seven different crossing angles, ranging from 0 to 180 degrees with intervals of 30 degrees.

The data provided consists of time-dependent trajectories of the participants who volunteered for the experiments. On day 1, there were 38 participants, and on day 2, there were 36 participants. Movement of participants were recorded using a VICON motion sensing infrared camera at a frequency of 120 Hz. Therefore, one time-step is equivalent to 1/120 seconds.

1.2 AIM OF THE PROJECT

The aim of the project is to analyze a real dataset of pedestrian crowd crossing flows collected in Rennes, France in 2016. The main objectives are to find the observed crossing angle for each trial, calculate velocities of all pedestrians along their trajectories, and define δ as a measure of deviation from the expected direction of motion. The project aims to plot normalized velocity and δ distributions for each crossing angle and investigate their dependence on α. Additionally, the project aims to summarize findings on the observed crossing angles and comment on the nature of the distribution curves.

The project is aimed at understanding the dynamic behavior of pedestrian crowds to provide effective methods for crowd management, given the increasing rate of urbanization and the life-threatening disasters resulting from overcrowding.

2 METHODOLOGY

2.1 FIRST STEPS – READING FILES

The first step needed to start the project was to determine which CSV file was for which theoretical angles. It is known that the trials were performed for seven different crossing angles, ranging from 0 to 180 degrees with intervals of 30 degrees. In order to notice the first differences between the files, I decided to write code that helped me read all the files from the selected folder in turn and see the "routes" of individual pedestrians.

It is known that in the data files, every 2i-th column corresponds to the x-position (in mm) of the i-th participant and every (2i + 1)-th column corresponds to its y-position (in mm), where i = 1, 2, 3, ..., 36 or, 38. To this end, I specified the number of pedestrians in a loop for each file in the folder, in order to read all the x- and y-positions of each participant in the study in the next loop. Then using the *plot* function I saw the route of each pedestrian.

To determine from which side the passers-by start the route and where they end it - using the *scatter* function, I marked their starting and ending positions - Figure 1.

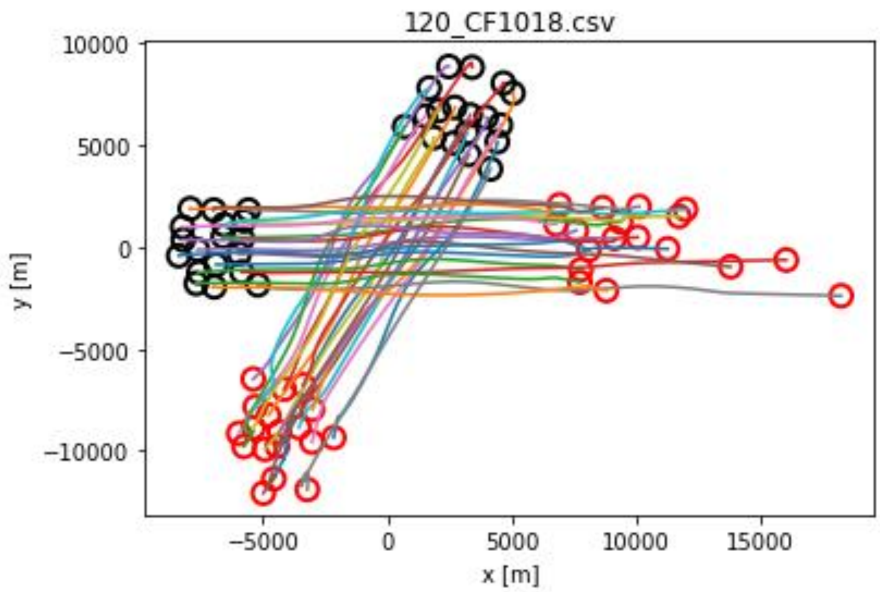

Figure 1. Example of plotting routes of pedestrians based on data in CF1018.csv file. Black points are refer to starting positions, while red ones are final positions.

Then, after plotting each file, I already knew which files belonged to which theoretical angles. I called the theoretical angles from 0 to 180 degrees with intervals of 30 degrees, that is:: 0, 30, 60, 90, 120, 150 and 180 [degrees] – which is showed in Table 1.

Table 1. Theoretical angles for each file.

Angeles [˚]	**0**	**30**	**60**	**90**	**120**	**150**	**180**
Files names	*CF106*	*CF227*	*CF223*	*CF218*	*CF211*	*CF214*	*CF207*
	CF105	*CF226*	*CF222*	*CF217*	*CF210*	*CF213*	*CF206*
	CF104	*CF225*	*CF221*	*CF216*	*CF209*	*CF212*	*CF205*
	CF103	*CF224*	*CF220*	*CF203*	*CF208*	*CF2008*	*CF204*
	CF102	*CF109*	*CF2020*	*CF202*	*CF2011*	*CF2007*	*CF2004*
	CF101	*CF108*	*CF2019*	*CF2016*	*CF2010*	*CF2006*	*CF2003*
	CF1004	*CF107*	*CF2018*	*CF2015*	*CF2009*	*CF2005*	*CF2002*
	CF1003	*CF2024*	*CF2017*	*CF2014*	*CF130*	*CF136*	*CF2001*
	CF1002	*CF2023*	*CF118*	*CF2013*	*CF129*	*CF135*	*CF142*
	CF1001	*CF2022*	*CF117*	*CF201*	*CF128*	*CF134*	*CF141*
		CF2021	*CF116*	*CF124*	*CF127*	*CF133*	*CF140*
		CF112	*CF115*	*CF122*	*CF126*	*CF132*	*CF139*
		CF111	*CF114*	*CF121*	*CF125*	*CF131*	*CF138*
		CF110	*CF113*	*CF120*	*CF1020*	*CF1024*	*CF137*
		CF1008	*CF1012*	*CF119*	*CF1019*	*CF1023*	*CF1028*
		CF1007	*CF1011*	*CF1016*	*CF1018*	*CF1022*	*CF1026*
		CF1006	*CF1010*	*CF1015*	*CF1017*	*CF1021*	*CF1025*
		CF1005		*CF1014*			
				CF1013			

After grouping the files into individual groups, I renamed them to include, at the beginning, a number denoting the theoretical angle at which the pedestrian groups should pass each other.

2.2 GROUPING PEDESTRIANS INTO TWO GROUPS

The most important part of the project was the correct grouping of pedestrians into two groups, which were passing each other. To do this, I applied a different rule for each group of files (dependent on the theoretical angle). I observed some dependencies for the different groups, where the individual pedestrians are located in the starting and final positions.

In general, my solution was to focus on finding the distance traveled by a particular pedestrian between the starting and ending positions. Based on the distance traveled in both x and y axes, I was able to assign the passerby to the appropriate group. Example of such grouped data is illustrated by Figure 2.

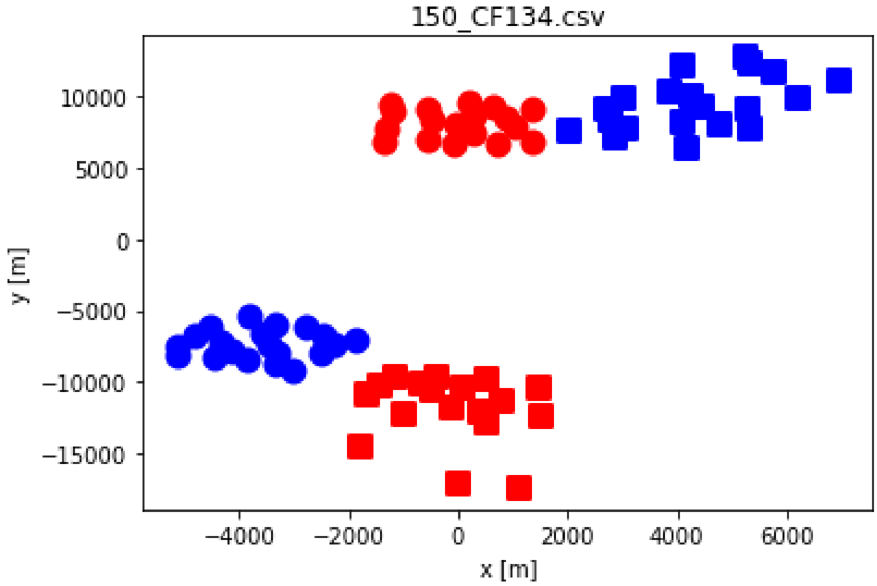

Figure 2. Data from file CF134.csv after grouping pedestrians into 2 groups: group 1 is colored with red, and group 2 is blue. Initial positions are marked as circles, finals positions are squares.

2.3 FINDING BARYCENTER IN EACH GROUP

The next key task was to find the barycenter in each group, and for both of the positions (initial and final). The implementation consisted of finding the " mean" position among all of the points. After determining each barycenter, I connected the found points to each other so that I could determine the actual direction of motion of a group in the form of lines. From the intersection of the lines for the two groups, I evaluated the observed crossing angle in each CSV file.

Below I would like to show how I obtained the observed crossing angle. First, I created two vectors based on four points (barycenters):

$$vector = (point_{2x} - point_{1x},\ \ point_{2y} - point_{1y})$$

Next I calculated the cosine of the angle between two vectors using the dot product of the two vectors, and then normalized the result by dividing by the product of their magnitudes.

Then I calculated the angle between two vectors in degrees using the *arccos* function, which takes the cosine of the angle as input and returns the angle in radians. The angle is then converted to degrees by multiplying by 180 and dividing by π.

The results of above actions are presented on Figure 3.

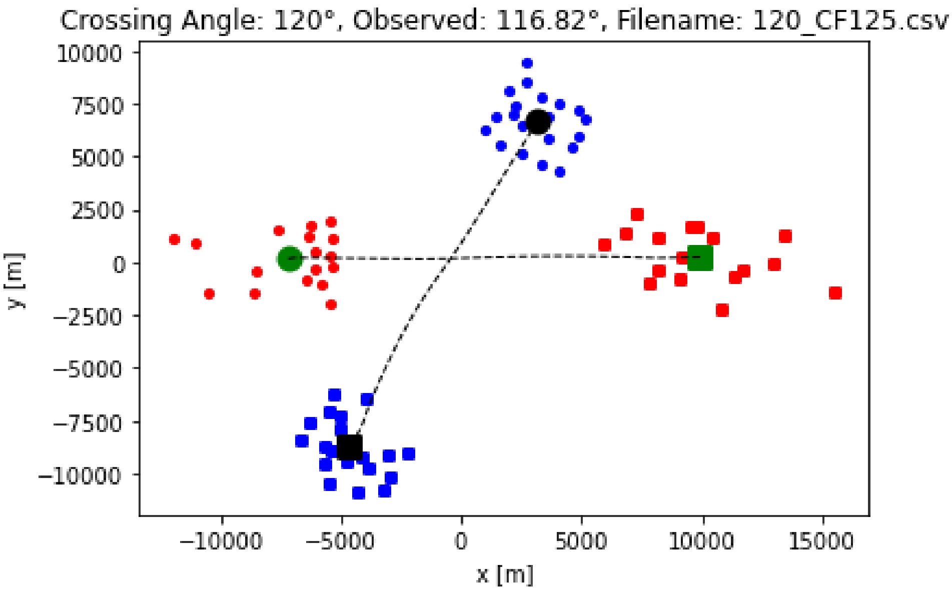

Figure 3. Data from file CF125.csv with barycenters, lines connected to them and observed angle in the title of the plot.

I grouped all observed angles into one table to see all differences.

2.4 VELOCITIES

The next task was to find the velocity for each pedestrian at each moment of time. It is known that one time-step is equivalent to 1/120 seconds. For this purpose, in a loop based on the input file folder, I calculated the number of pedestrians in a given dataset and then created new columns in the number of pedestrians stored in a given file. The next step was to calculate the individual positions - current and previous - and at each time step calculate the speed according to the formula:

$$v_i(t) = \left[\frac{\sqrt{(x_i(t) - x_i(t-1))^2 + (y_i(t) - y_i(t-1))^2}}{1000} \times 120\right] m/sec$$

where $[x_i(t), y_i(t)]$ denotes the position of the i-th agent at time t, as recorded in the data.

I saved the velocities in the previously created columns and saved the finished files in a folder created for this purpose. Then I plot normalized velocity distributions for each α considering all the data files.

2.5 DEVIATIONS

I used a similar approach to calculate δ. Delta (δ) is a measure of the deviations that a pedestrian makes with his/her originally instructed direction of motion. I calculated δ at every instant of a pedestrian's trajectory as the angle between the trajectory of that pedestrian and his/her expected direction of motion.

$$\sigma = actual\ direction - expected\ direction$$

The pedestrian's expected direction of motion it's simply angle between first and last point on both axes:

$$\sigma = \tan^{-1}\frac{y_i(t) - y_i(t-1)}{x_i(t) - x_i(t-1)} - \tan^{-1}\frac{y_{last} - y_{first}}{x_{last} - x_{first}}$$

where $[x_i(t), y_i(t)]$ denotes the position of the i-th agent at time t, as recorded in the data and $[x_{first}, y_{first}]$ is the position in 1st time step, $[x_{last}, y_{last}]$ is the position in the last time step.

I saved the deviations in the previously created columns and saved the finished files in a folder created for this purpose. Then I plot normalized distributions of δ for each α considering all the data files.

3 RESULTS AND DISCUSSIONS

For all the data files I have found the observed crossing angle. I summarized them in Table 2.

Table 2. Summary of observed angles for each of the data files for each of angles between two groups.

Angle between two groups	30	60	90	120	150	180
Observed angles	25,91735703	64,37966634	91,31522041	117,517378	153,8491476	179,8082259
	26,78309993	60,51967827	89,74545678	116,3146259	153,9820276	179,6765464
	25,33173928	63,31271765	90,35647202	116,821227	153,9936358	179,4851955
	25,09392164	63,94773563	89,44109071	116,8553866	154,0408676	179,8814409
	26,34448893	63,10955896	89,71390376	116,8160957	153,4921241	179,6248721
	27,7029854	63,72654607	89,91040646	115,8661246	155,0298129	179,3265277
	26,13294028	64,37408079	90,6475714	118,1457956	154,4734532	179,6798647
	26,13185035	65,07670339	90,47904192	115,4143668	155,0329797	179,706347
	26,0479385	60,09307231	90,11460248	116,8551579	153,6825404	179,8099231
	25,58706119	63,62203297	88,43150953	116,1377194	154,2777183	179,667641
	27,79161832	65,09560281	89,5193589	116,1405381	154,7469821	179,8380737
	26,07937102	63,47310016	90,70628846	117,0722698	154,3002488	179,8793485
	26,4034677	63,99996751	91,01858261	116,8365452	153,2133231	179,9391156
	26,46724016	62,33666067	90,23405892	115,9500186	153,8459784	179,916502
	26,88639001	64,7257799	89,47256377	115,6756918	153,6407417	179,9142583
	25,39262949	63,73079348	89,85632765	117,9254714	154,6756371	179,3856059
	26,04227641	63,19759073	89,78313385	116,9062586	154,5654979	179,0861238
	25,20395842		90,95821195			
			90,46284572			
Mean	**26,19**	**63,45**	**90,11**	**116,66**	**154,17**	**179,68**

Also, I calculated mean to find out how much the obtained results differ from the theoretical assumptions. The observed results actually differ from the theoretical ones, but the differences are not large. The largest were observed with groups crossing at 150 and 30 degrees. The groups found it easiest to maintain the assumed angle when they passed each other at 180 and 90 degrees.

Then I calculated velocities of all the pedestrians at all instances along their trajectories. The following Figure 4 represents normalized velocity distributions for the pedestrians corresponding to each α considering all the data files.

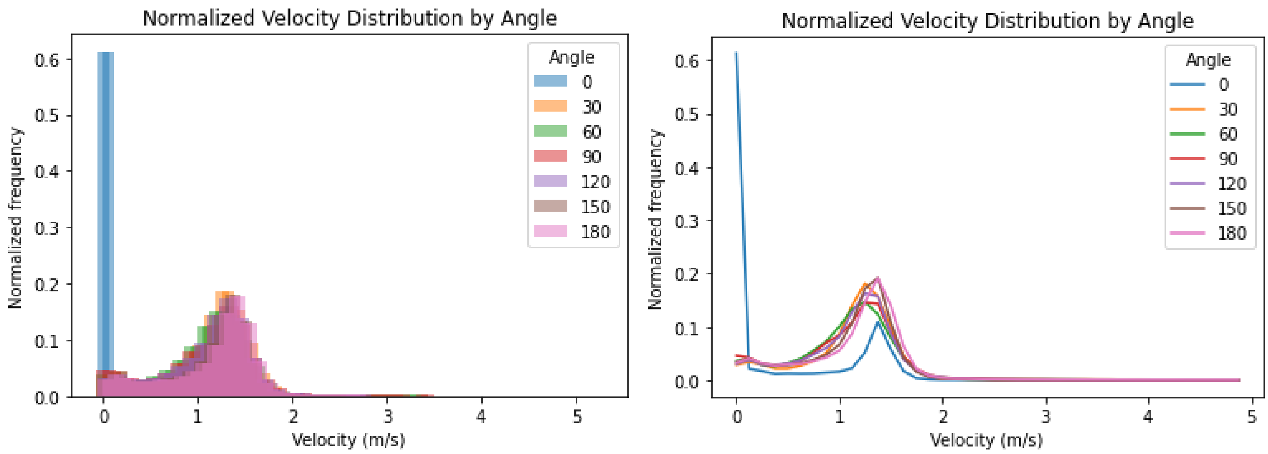

Figure 4. Normalized velocity distributions for the pedestrians corresponding to each α considering all the data files. Graph on the left is generated as histogram, graph on the right is linear.

The distribution curves clearly depend on the angle α. It can be easily seen on Figure 5. For an angle of 0 degrees, the velocities of the pedestrians are very low, as most of them are just above 0 m/s. The reason is overcrowding of space – pedestrian doesn't have free way to go in the direction that he want.

The highest number of counts for velocities in the range 0.8-1.5 m/s was obtained for groups crossing at an angle of 30 degrees. In that case pedestrians have the greatest ability to move at their desired speed.

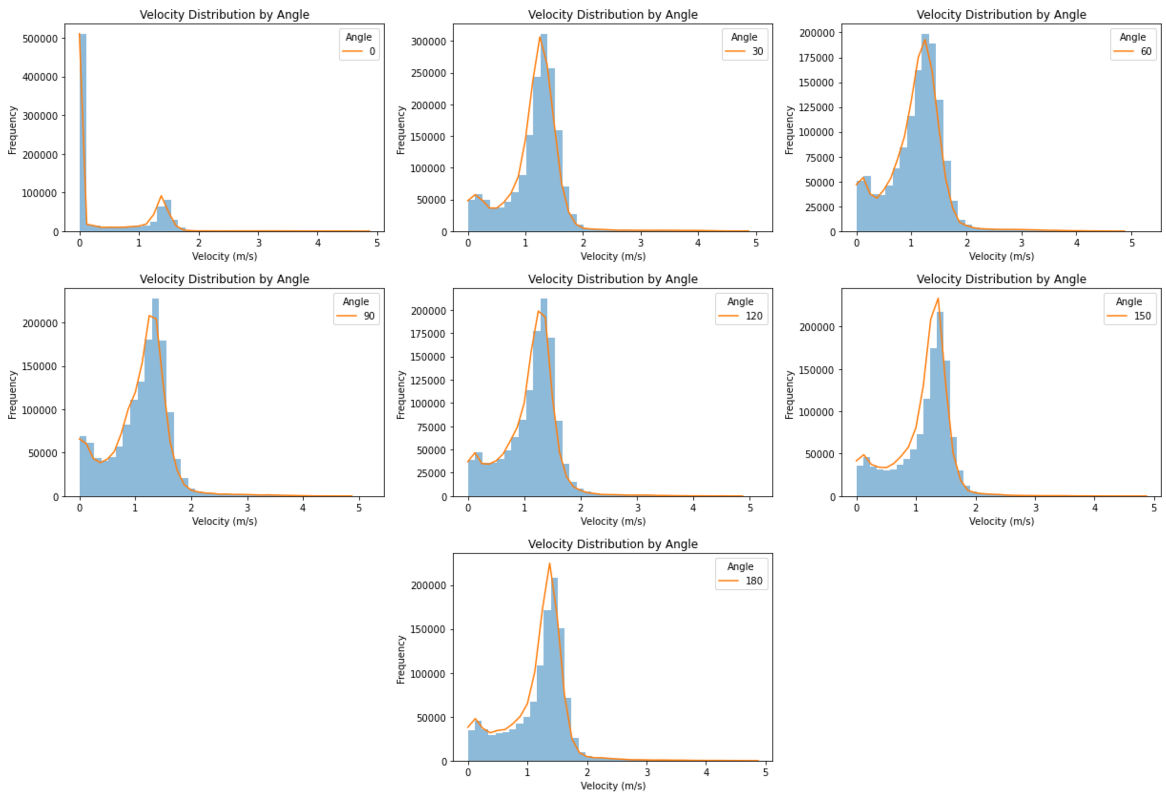

Figure 5. Velocity distributions for the pedestrians corresponding to each α considering all the data files, showed in separate graphs for each angle.

The last task was to calculate δ at every instant of a pedestrian's trajectory as the angle between the trajectory of that pedestrian and his/her expected direction of motion. Figure 6 represents normalized deviation distributions for the pedestrians corresponding to each α considering all the data files.

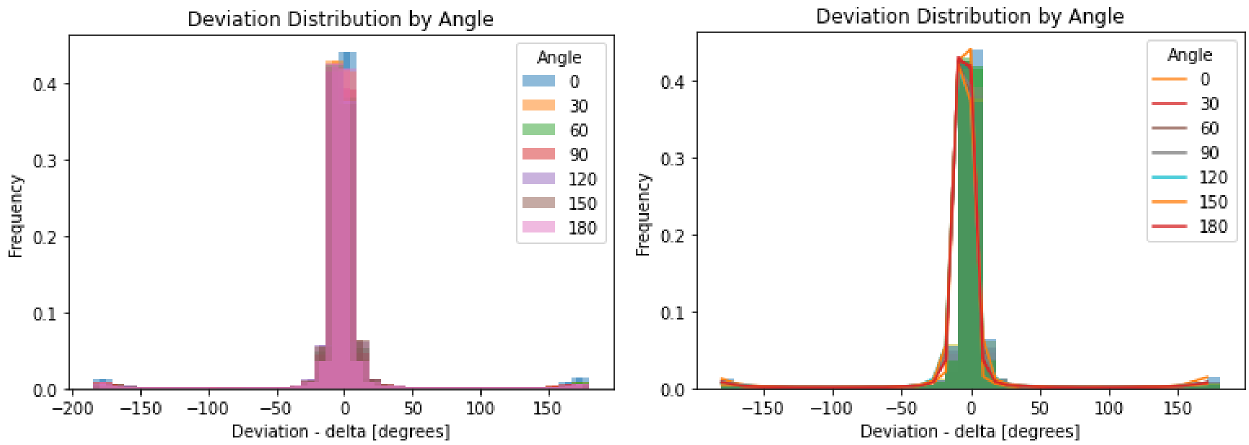

Figure 6. Normalized deviation distributions for the pedestrians corresponding to each α considering all the data files.

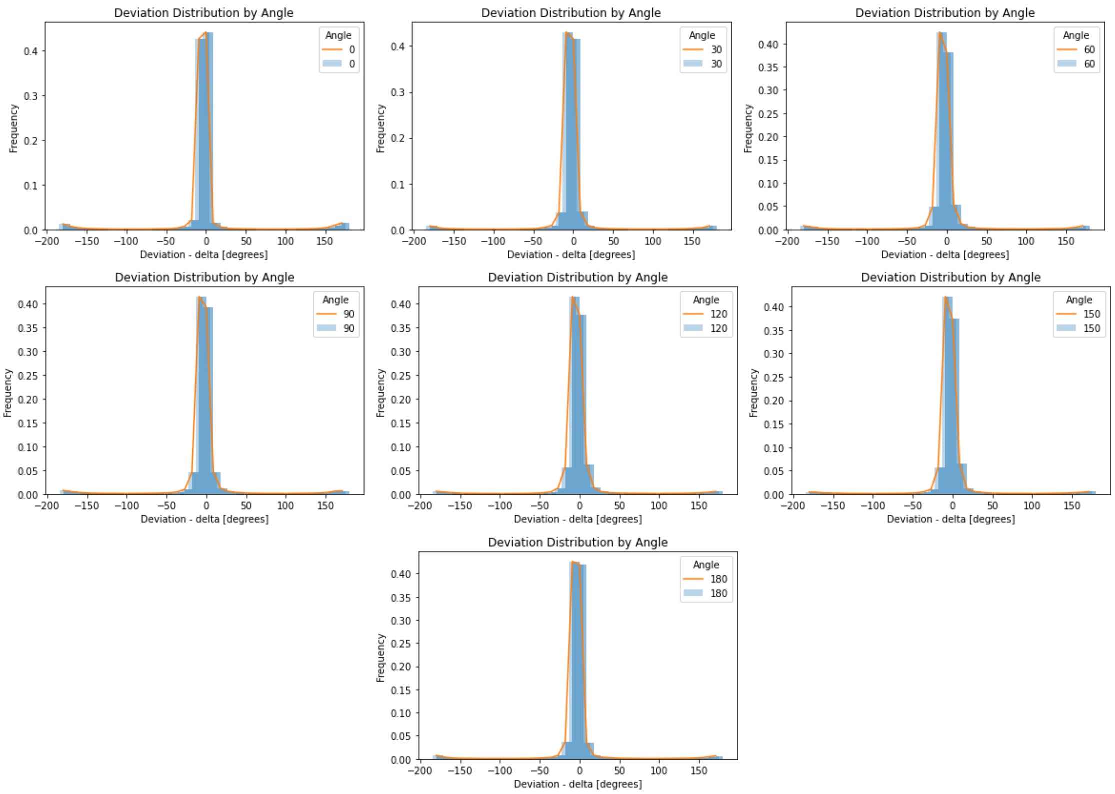

Figure 7. Normalized deviation distributions for the pedestrians corresponding to each α considering all the data files, showed in separate graphs for each angle.

The distribution curves also depend on the angle α. But it is less visible than in case of velocities. Generally, the frequency of occurrences of small angles (around 0 degrees) is the highest.

An interesting phenomenon is that in some cases (this can be seen well for the angle of crossing groups of 0 degrees) there is a noticeable number of occurrences when the pedestrians turn back (deviations were -180 or 180 degrees). It can be mostly seen it on Figure 7.

4 CONCLUSION

In conclusion, the project aimed to analyze a real dataset of pedestrian crowd crossing flows to understand the dynamic behavior of pedestrian crowds for effective crowd management. The data was collected in Rennes, France in 2016, and the main objective was to find the observed crossing angle for each trial, calculate velocities of all pedestrians along their trajectories, and define δ as a measure of deviation from the expected direction of motion.

The project aimed to plot normalized velocity and δ distributions for each crossing angle and investigate their dependence on α. Additionally, the project aimed to summarize findings on the observed crossing angles and comment on the nature of the distribution curves. The methodology involved reading files, grouping pedestrians into two groups, finding the barycenter in each group, and calculating the observed crossing angle using vector calculations to get velocity and deviation of each pedestrian.

After finding the observed crossing angle for each trial, the next step was to calculate the velocities of all pedestrians along their trajectories. This was done by computing the distance between consecutive time steps for each pedestrian and dividing it by the time interval (1/120 seconds). This gave us velocity of each pedestrian at each time step.

Next I defined δ as a measure of deviation from the expected direction of motion. The expected direction of motion for each pedestrian was defined as the line connecting the initial and final position of each pedestrian. δ was defined as the angle between the velocity vector of each pedestrian and the expected direction of motion.

The analysis of the distributions showed that the normalized velocity and deviation distribution was nearly Gaussian for all crossing angles.

The results show that the distributions of velocity and deviation are dependent on the crossing angle and highlight the need for further research in this area. The distributions of velocity and deviation of pedestrians can be dependent on the crossing angle because crossing angle affects the perceived safety and comfort of pedestrians while crossing a street. At some angles pedestrians have better visibility of oncoming traffic, and they may feel more comfortable crossing the street at a faster speed. In the case of a 0-degree angle, all pedestrians in both groups are walking in one direction, which results in a reduction of space among people and at the same time slows them down.

Overall, the project provides valuable insights into crowd behavior and can help in developing effective crowd management strategies to prevent life-threatening disasters resulting from overcrowding.