PROJECT TITLE

DATE

Project Authors	Team OK: Ouyang Du, Kara Wei	**Status**	On target/ at risk
Industrial Area of Application	Garment Industry, Labor Market		

Project Description / Deliverables

[summary of your project and deliverables]

The main interest of our project is Employee Productivity. By understanding the variables that impact employee efficiency, businesses can make more informed decisions on ways they can help their workers work better. For this project, we will first look at distributions of variable to deal with potential dataset imbalance. After cleaning our dataset, we will look at the assumptions required for model building. Our final task is to build a plausible model that can predict productivity of workers effectively.

The dataset we use contain 1197 entries and 15 variables. The dataset we plan on using is the Productivity Prediction of Garment Employees. Based on manufacturing companies, this dataset provides information on Garment workers' performance from different departments. The dataset contains a total of 15 variables: date, day of the week, quarter, department, number of teams, number of works, number of style changes, targeted productivity, Standard Minute Value (SMV), Work in Progress (WIP), overtime per team in minutes, amount of financial incentives, amount of time when production was interrupted, number of works who were idle, and actual productivity (" Productivity Prediction of Garment Employees").

Reference:

" Productivity Prediction of Garment Employees." *Archive.ics.uci.edu*, UCI Machine Learning Repository, archive.ics.uci.edu/dataset/597/productivity+prediction+of+garment+employees. Accessed 8 Sept. 2023.

finnstats. *Imputing Missing Values in R | R-Bloggers*. 9 Mar. 2022, www.r-bloggers.com/2022/03/imputing-missing-values-in-r/.

avcontentteam. "R Packages | Impute Missing Values in R." Analytics Vidhya, 4 Mar. 2016, www.analyticsvidhya.com/blog/2016/03/tutorial-powerful-packages-imputing-missing-values/.

"Multivariate Imputation by Chained Equations." *R-Packages*, cran.r-project.org/web/packages/mice/readme/README.html. Accessed 4 Oct. 2023.

Decisions Impacted

- **What variabless are associated with fluctuation of productivity?**
- **Should we anticipate an upward or downward trend in productivity for the future?**

Expected Business Value

- **[what is the business value of your project]**
- By studying and modeling the relationship between goal setting and productivity, business managers can have more information on how they can set goals based on the data that is available to them. From a societal view, we can study the longitudinal effects of productivity. Will a large sustained workload affect the productivity of the following quarters? The answer to this question would help businesses avoid employee burnout, which will increase employee satisfaction.

Weekly Progress

- **Describe activities performed during week towards your milestones**
- We started by developing histograms for all numerical variables: 1) Productivity Index, 2) Target Productivity, 3) Standard Minute Value, 4) Work in Progress, 5) Over Time, 6) Financial Incentive, 7) Idle Time, 8) Idle Workers, 9) Number of Style Changed, and 10) Number of Workers in a Team.
- After observing the distribution, our main task was to deal with the missing values. We noticed that the variable wip has almost half of the values as NA.
- We then implemented multiple methods: 1) remove rows with missing values, 2) replacing with zeros, 3) replacing with mean, 4) replacing with median, 5) replacing with predictive values after regressing wip to other variables (MICE package).
- For each method, we used the histogram of wip as reference for validity purpose. We looked into whether the distribution is biased: 1) is the model bell shaped? 2) are there heavy tails presented?
- We communicated frequently to make sure we are on track. We developed our plan for next week.
- **Link to project repository:** https://github.com/OuyangDu/OK.git
- **Dataset:** https://www.kaggle.com/datasets/ishadss/productivity-prediction-of-garment-employees
- Screenshot of histograms (codes in repo)
- screenshot of cleaned dataset (head):

	date	quarter	department	day	team	targeted_productivity	smv	wip
	<chr>	<chr>	<chr>	<chr>	<int>	<dbl>	<dbl>	<dbl>
1	1/1/2015	Quarter1	sweing	Thursday	8	0.80	26.16	1108.000
2	1/1/2015	Quarter1	finishing	Thursday	1	0.75	3.94	1190.466
3	1/1/2015	Quarter1	sweing	Thursday	11	0.80	11.41	968.000
4	1/1/2015	Quarter1	sweing	Thursday	12	0.80	11.41	968.000
5	1/1/2015	Quarter1	sweing	Thursday	6	0.80	25.90	1170.000
6	1/1/2015	Quarter1	sweing	Thursday	7	0.80	25.90	984.000