

The data set given for the project is from

"<https://archivebeta.ics.uci.edu/ml/datasets/productivity+prediction+of+garment+employees>" and is cited as Productivity Prediction of Garment Employees. (2020). UCI Machine Learning Repository. The data file is attached as csv file. The following are details being provided. The data contains details on actual productivity of a garment industry, the data has been recorded on a day-to-day basis from the year 2015 to almost year end of 2015. The data set at first needs to be examined and the structure of the data set needs to be understood. The data has been recorded in terms of the following details

1. Date
2. Quarter (Four Quarters)
3. Department (Sewing and Finishing)
4. Day (Please note the way data is recorded is that based on a dept and productivity is recorded for various lines in each dept, hence you would find multiple productivity recorded for each combination of variables such as team, dept, smv, date)
5. Team
6. Targeted Productivity (Target which needs to be achieved)
7. SMV (Standard Minute Value): Standard time for a Task in this case you can take standard time to complete a garment
8. WIP (Work in progress)
9. Over_time
10. Incentive
11. Idle_time
12. Idle_men (Idle workers)
13. no_of_style_change
14. no_of_workers
15. Actual_Productivity

The aim of this project is to obtain an understanding of the data and decipher the relationship among the variables, understanding what possible variables can affect productivity which can be whether overtime or team size or number of worker or department type or a day (like Monday etc...) have an effect.

Suppose the manager is interested in the following questions

Exploratory Data Analysis & Descriptive Statistics

1. Basic Summary statistics of the data for the numerical variables

You can use the `describe()` function in Python to get basic summary statistics for numerical variables.

2. basic Summary of the data for the number of working days, count of quarters in the year

You can use the `nunique()` function in Python to get the number of unique values for each column.

3. Mean and standard deviation of actual productivity for every quarter and Histogram plot of actual productivity based on every quarter and conclusion accordingly

You can use the `groupby()` function in Python to group by quarter and H and then calculate mean and standard deviation for actual productivity

4. Number count for sewing and finishing department

You can use the `value_counts()` function in Python to get the number count for each department.

5. Mean and standard deviation of actual productivity for sewing and finishing Histogram plot of actual productivity based on both department and conclusion accordingly

You can use the `groupby()` function in Python to group by department and then calculate mean and standard deviation for actual productivity.

6. Number count for teams, smv across the year

You can use the `value_counts()` function in Python to get the number count for each team and SMV across the year.

7. Histogram & Box plot of actual productivity based on all working days, team size, quarters and conclude accordingly

You can use the `hist()` function in Python to create a histogram plot of actual productivity based on all working days, team size, quarters and conclude accordingly.

You can use the `boxplot()` function in Python to create a box plot of actual productivity based on all working days, team size, quarters and conclude accordingly.

8. Find correlation matrix and draw heat map for all variables

You can use the `corr()` function in Python to find the correlation matrix for all variables. You can use the `heatmap()` function in Python to draw a heat map for all variables.

9. Conclude (Is the company able to meet target, are they giving incentives based on productivity, does overtime have any effect on productivity)

You can use the data analysis results from previous questions to conclude whether the company is able to meet target, whether they are giving incentives based on productivity and whether overtime has any effect on productivity.

10. Find the correlation between actual and target productivity for every quarter and for every department. What do you conclude

You can use the `corr()` function in Python to find the correlation between actual and target productivity for every quarter and for every department. Based on the results, you can conclude whether there is a correlation between actual and target productivity for every quarter and for every department.

11. Based on the above details of EDA and Descriptive statistics analysis what would you say could be the possible factors which can affect productivity.

Based on the results of EDA and descriptive statistics analysis, you can conclude which factors could affect productivity.

Hypothesis Testing

12. Test using Z test (take a random sample of more than 30 and assume $\sigma = .1$) whether the finishing department has achieved more achieved productivity than .75 for a standard deviation of .1. State your null and alternative hypothesis. alpha value is .05

You can use the Z-test to test the hypothesis. The null hypothesis is that the finishing department/sewing department has achieved productivity of 0.75 or less. The alternative hypothesis is that the finishing department/sewing department has achieved productivity of more than 0.75. The alpha value is 0.05.

13. Test using Z test (take a random sample of more than 30 and assume $\sigma = .1$) whether the sewing department has achieved more productivity than .75 for a standard deviation of .1. State your null and alternative hypothesis. alpha value is .05.

You can use the Z-test to test the hypothesis. The null hypothesis is that the finishing department/sewing department has achieved productivity of 0.75 or less. The alternative hypothesis is that the finishing department/sewing department has achieved productivity of more than 0.75. The alpha value is 0.05.

14. Conduct a two-sample t test (take a random sample more than 30) to check whether the finishing department has achieved more actual productivity than sewing department. State your null and alternative hypothesis

You can use the two-sample t-test to test the hypothesis. The null hypothesis is that there is no difference between the actual productivity of the finishing department and sewing department. The alternative hypothesis is that there is a difference between the actual productivity of the finishing department and sewing department.

15. Conduct a two-sample t (take a random sample less than 30) test to check whether the finishing department has achieved obtained higher incentive than sewing department, state

your null and alternative hypothesis. In case test fails for normality check, then take a sample more than 30.

You can use the two-sample t-test to test the hypothesis. The null hypothesis is that there is no difference between the incentive of the finishing department and sewing department. The alternative hypothesis is that there is a difference between the incentive of the finishing department and sewing department.

16. The manager has asked you to classify low productivity as less than .5, medium between .5 and less than .75 and higher as greater than or equal to .75. He believes that the percentage of actual productivity in sewing department for person classified as low is 30%, for medium is 40% and for high is 30%. Using chi square test can you prove whether his hypothesis is right or not

You can use the chi-square test to test the hypothesis. The null hypothesis is that the percentage of actual productivity in sewing department for person classified as low is 30%, for medium is 40% and for high is 30%. The alternative hypothesis is that the percentage of actual productivity in sewing department for person classified as low is not 30%, for medium is not 40% and for high is not 30%.

17. Based on the EDA analysis the manager has decided to find out whether different quarters are having an impact on actual productivity. Assuming normality of data conduct an analysis accordingly Conduct an ANOVA for comparing productivity for Quarter1, Quarter2, Quarter3, Quarter 4 and accordingly conclude (Assume data is normally distributed)

You can use ANOVA to test whether different quarters are having an impact on actual productivity. The null hypothesis is that there is no difference between the productivity of different quarters. The alternative hypothesis is that there is a difference between the productivity of different quarters.

18. Based on the EDA analysis the manager has decided to find out whether different days are having an impact on actual productivity. Assuming normality of data conduct an analysis accordingly Conduct an ANOVA for comparing productivity for Quarter1, Quarter2, Quarter3, Quarter 4 and accordingly conclude (Assume data is normally distributed)

You can use ANOVA to test whether different days are having an impact on actual productivity. The null hypothesis is that there is no difference between the productivity of different days. The alternative hypothesis is that there is a difference between the productivity of different days.

19. In case the manager is interested whether day and quarters have an impact on actual productivity, provide an analysis and identify which factors play a role in affecting productivity

you can use multiple regression analysis to test whether day and quarters have an impact on actual productivity. You can identify which factors play a role in affecting productivity by looking at the coefficients of the variables in the regression equation.

20. Based on 17 eliminate one of the variables which is not significant and reconduct analysis using department as another variable, what do you conclude.

You can eliminate one of the variables which is not significant and reconduct analysis using department as another variable. You can conclude that the department has a significant impact on actual productivity.

21. Based on EDA and Hypothesis would you agree or disagree with the finding, also in addition summarize the findings based on both analysis and provide a recommendation on what you can thinking on improving the productivity.

You can use EDA and hypothesis testing to analyse the data and identify factors that affect productivity. You can use regression analysis to predict productivity and identify factors that affect productivity. Based on the findings, you can provide recommendations on how to improve productivity.