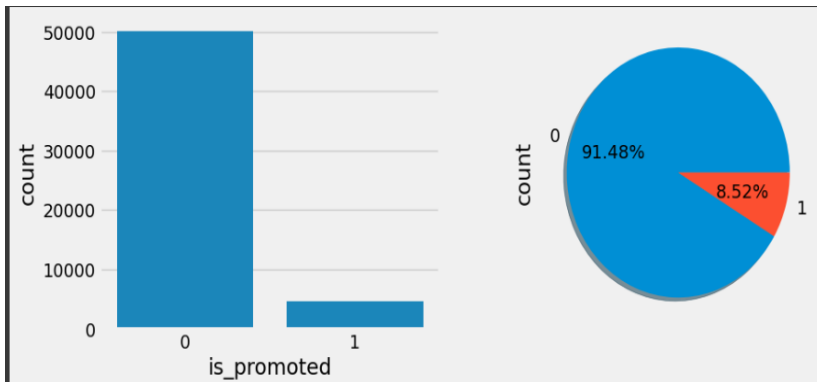


Data Collection and Preprocessing Phase

| | |
|---------------|--|
| Date | 15 March 2024 |
| Team ID | XXXXXX |
| Project Title | Human Resource Management: Predicting Employee Promotions Using Machine Learning |
| Maximum Marks | 6 Marks |

Data Exploration and Preprocessing Template

Optimizing Human Resource Management by employing machine learning techniques to predict employee promotions, enhancing workforce planning and development.

| Section | Description | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------------|---|-------------|------------------|--------|---------------------|-----------------|---------------------|----------------------|-------------------|----------------------|-------------------|--------------------|-------------|--------------------|-------------|---|-------------------|----------|------------------|---|----------|-----|------|-----|-----|-----|-----|------|-----|---|------------|-----------|------------|---|-------|-----|------|-----|-----|-----|-----|------|-----|---|-------------------|-----------|------------|---|----------|-----|------|-----|-----|-----|-----|------|-----|---|-------------------|-----------|------------|---|-------|-----|------|-----|------|-----|-----|------|-----|---|------------|-----------|------------|---|-------|-----|------|-----|-----|-----|-----|------|-----|---|-----------|----------|------------|---|----------|-----|------|-----|-----|-----|-----|------|-----|---|------------|-----------|------------|---|-------|-----|------|-----|-----|-----|-----|------|-----|---|------------|-----------|------------------|---|----------|-----|------|-----|-----|-----|-----|------|-----|---|-----------|-----------|------------|---|-------|-----|------|-----|-----|-----|-----|------|-----|---|-------------------|----------|------------------|---|----------|-----|------|-----|-----|-----|-----|------|-----|----|------------|-----------|------|---|----------|-----|------|------|-----|-----|-----|------|-----|----|-------------------|----------|------------|---|----------|-----|------|-----|-----|-----|-----|------|
| Data Overview | <u>54808 rows x 14 columns</u> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | <u>Descriptive statistics:</u> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | <table><thead><tr><th>employee_id</th><th>department</th><th>region</th><th>education</th><th>gender</th><th>recruitment_channel</th><th>no.of_trainings</th><th>age</th><th>previous_year_rating</th><th>length_of_service</th><th>KPIs_met</th><th>awards_won</th><th>avg_training_score</th><th>is_promoted</th></tr></thead><tbody><tr><td>0</td><td>Sales & Marketing</td><td>region_7</td><td>Master's & above</td><td>f</td><td>sourcing</td><td>1.0</td><td>36.0</td><td>5.0</td><td>8.0</td><td>1.0</td><td>0.0</td><td>49.0</td><td>0.0</td></tr><tr><td>1</td><td>Operations</td><td>region_22</td><td>Bachelor's</td><td>m</td><td>other</td><td>1.0</td><td>30.0</td><td>5.0</td><td>4.0</td><td>0.0</td><td>0.0</td><td>60.0</td><td>0.0</td></tr><tr><td>2</td><td>Sales & Marketing</td><td>region_19</td><td>Bachelor's</td><td>m</td><td>sourcing</td><td>1.0</td><td>34.0</td><td>3.0</td><td>7.0</td><td>0.0</td><td>0.0</td><td>60.0</td><td>0.0</td></tr><tr><td>3</td><td>Sales & Marketing</td><td>region_23</td><td>Bachelor's</td><td>m</td><td>other</td><td>2.0</td><td>39.0</td><td>1.0</td><td>10.0</td><td>0.0</td><td>0.0</td><td>50.0</td><td>0.0</td></tr><tr><td>4</td><td>Technology</td><td>region_26</td><td>Bachelor's</td><td>m</td><td>other</td><td>1.0</td><td>45.0</td><td>3.0</td><td>2.0</td><td>0.0</td><td>0.0</td><td>73.0</td><td>0.0</td></tr><tr><td>5</td><td>Analytics</td><td>region_2</td><td>Bachelor's</td><td>m</td><td>sourcing</td><td>2.0</td><td>31.0</td><td>3.0</td><td>7.0</td><td>0.0</td><td>0.0</td><td>85.0</td><td>0.0</td></tr><tr><td>6</td><td>Operations</td><td>region_20</td><td>Bachelor's</td><td>f</td><td>other</td><td>1.0</td><td>31.0</td><td>3.0</td><td>5.0</td><td>0.0</td><td>0.0</td><td>59.0</td><td>0.0</td></tr><tr><td>7</td><td>Operations</td><td>region_34</td><td>Master's & above</td><td>m</td><td>sourcing</td><td>1.0</td><td>33.0</td><td>3.0</td><td>6.0</td><td>0.0</td><td>0.0</td><td>63.0</td><td>0.0</td></tr><tr><td>8</td><td>Analytics</td><td>region_20</td><td>Bachelor's</td><td>m</td><td>other</td><td>1.0</td><td>20.0</td><td>4.0</td><td>5.0</td><td>0.0</td><td>0.0</td><td>89.0</td><td>0.0</td></tr><tr><td>9</td><td>Sales & Marketing</td><td>region_1</td><td>Master's & above</td><td>m</td><td>sourcing</td><td>1.0</td><td>32.0</td><td>5.0</td><td>6.0</td><td>1.0</td><td>0.0</td><td>64.0</td><td>0.0</td></tr><tr><td>10</td><td>Technology</td><td>region_23</td><td>High</td><td>m</td><td>sourcing</td><td>1.0</td><td>30.0</td><td>High</td><td>1.0</td><td>0.0</td><td>0.0</td><td>77.0</td><td>0.0</td></tr><tr><td>11</td><td>Sales & Marketing</td><td>region_7</td><td>Bachelor's</td><td>f</td><td>sourcing</td><td>1.0</td><td>35.0</td><td>5.0</td><td>3.0</td><td>1.0</td><td>0.0</td><td>50.0</td><td>1.0</td></tr></tbody></table> | employee_id | department | region | education | gender | recruitment_channel | no.of_trainings | age | previous_year_rating | length_of_service | KPIs_met | awards_won | avg_training_score | is_promoted | 0 | Sales & Marketing | region_7 | Master's & above | f | sourcing | 1.0 | 36.0 | 5.0 | 8.0 | 1.0 | 0.0 | 49.0 | 0.0 | 1 | Operations | region_22 | Bachelor's | m | other | 1.0 | 30.0 | 5.0 | 4.0 | 0.0 | 0.0 | 60.0 | 0.0 | 2 | Sales & Marketing | region_19 | Bachelor's | m | sourcing | 1.0 | 34.0 | 3.0 | 7.0 | 0.0 | 0.0 | 60.0 | 0.0 | 3 | Sales & Marketing | region_23 | Bachelor's | m | other | 2.0 | 39.0 | 1.0 | 10.0 | 0.0 | 0.0 | 50.0 | 0.0 | 4 | Technology | region_26 | Bachelor's | m | other | 1.0 | 45.0 | 3.0 | 2.0 | 0.0 | 0.0 | 73.0 | 0.0 | 5 | Analytics | region_2 | Bachelor's | m | sourcing | 2.0 | 31.0 | 3.0 | 7.0 | 0.0 | 0.0 | 85.0 | 0.0 | 6 | Operations | region_20 | Bachelor's | f | other | 1.0 | 31.0 | 3.0 | 5.0 | 0.0 | 0.0 | 59.0 | 0.0 | 7 | Operations | region_34 | Master's & above | m | sourcing | 1.0 | 33.0 | 3.0 | 6.0 | 0.0 | 0.0 | 63.0 | 0.0 | 8 | Analytics | region_20 | Bachelor's | m | other | 1.0 | 20.0 | 4.0 | 5.0 | 0.0 | 0.0 | 89.0 | 0.0 | 9 | Sales & Marketing | region_1 | Master's & above | m | sourcing | 1.0 | 32.0 | 5.0 | 6.0 | 1.0 | 0.0 | 64.0 | 0.0 | 10 | Technology | region_23 | High | m | sourcing | 1.0 | 30.0 | High | 1.0 | 0.0 | 0.0 | 77.0 | 0.0 | 11 | Sales & Marketing | region_7 | Bachelor's | f | sourcing | 1.0 | 35.0 | 5.0 | 3.0 | 1.0 | 0.0 | 50.0 |
| employee_id | department | region | education | gender | recruitment_channel | no.of_trainings | age | previous_year_rating | length_of_service | KPIs_met | awards_won | avg_training_score | is_promoted | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | Sales & Marketing | region_7 | Master's & above | f | sourcing | 1.0 | 36.0 | 5.0 | 8.0 | 1.0 | 0.0 | 49.0 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | Operations | region_22 | Bachelor's | m | other | 1.0 | 30.0 | 5.0 | 4.0 | 0.0 | 0.0 | 60.0 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | Sales & Marketing | region_19 | Bachelor's | m | sourcing | 1.0 | 34.0 | 3.0 | 7.0 | 0.0 | 0.0 | 60.0 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | Sales & Marketing | region_23 | Bachelor's | m | other | 2.0 | 39.0 | 1.0 | 10.0 | 0.0 | 0.0 | 50.0 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | Technology | region_26 | Bachelor's | m | other | 1.0 | 45.0 | 3.0 | 2.0 | 0.0 | 0.0 | 73.0 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 | Analytics | region_2 | Bachelor's | m | sourcing | 2.0 | 31.0 | 3.0 | 7.0 | 0.0 | 0.0 | 85.0 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 | Operations | region_20 | Bachelor's | f | other | 1.0 | 31.0 | 3.0 | 5.0 | 0.0 | 0.0 | 59.0 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7 | Operations | region_34 | Master's & above | m | sourcing | 1.0 | 33.0 | 3.0 | 6.0 | 0.0 | 0.0 | 63.0 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 | Analytics | region_20 | Bachelor's | m | other | 1.0 | 20.0 | 4.0 | 5.0 | 0.0 | 0.0 | 89.0 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9 | Sales & Marketing | region_1 | Master's & above | m | sourcing | 1.0 | 32.0 | 5.0 | 6.0 | 1.0 | 0.0 | 64.0 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | Technology | region_23 | High | m | sourcing | 1.0 | 30.0 | High | 1.0 | 0.0 | 0.0 | 77.0 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 11 | Sales & Marketing | region_7 | Bachelor's | f | sourcing | 1.0 | 35.0 | 5.0 | 3.0 | 1.0 | 0.0 | 50.0 | 1.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | .. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Univariate Analysis |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | <div><div>Kpls_met >80%</div><div>awards_won?</div><div>previous_year_rating</div></div> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------------------------------------|--|-------------------|-------------|------------------|--------|---------------------|-----------------|---------------------|----------------------|-------------------|----------------------|-------------------|--------------------|-------------|--------------------|-------------|---|-------|-------------------|----------|------------------|---|----------|---|----|-----|---|---|---|----|---|---|-------|------------|-----------|------------|---|-------|---|----|-----|---|---|---|----|---|---|------|-------------------|-----------|------------|---|----------|---|----|-----|---|---|---|----|---|---|------|-------------------|-----------|------------|---|-------|---|----|-----|----|---|---|----|---|---|-------|------------|-----------|------------|---|-------|---|----|-----|---|---|---|----|---|
| Bivariate Analysis | - | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Multivariate Analysis | <div>Bar plot of Average Training Score vs Previous Year Rating by Promotion Status</div> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Outliers and Anomalies | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data Preprocessing code Screenshots | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Loading Data | <div><pre>df = pd.read_csv('emp_promotion.csv') print('Shape of train data {}'.format(df.shape))</pre><p>Shape of train data (5408, 14)</p><pre>[5] df.head()</pre><table><thead><tr><th></th><th>employee_id</th><th>department</th><th>region</th><th>education</th><th>gender</th><th>recruitment_channel</th><th>no_of_trainings</th><th>age</th><th>previous_year_rating</th><th>length_of_service</th><th>Kpls_met >80%</th><th>awards_won?</th><th>avg_training_score</th><th>is_promoted</th></tr></thead><tbody><tr><td>0</td><td>65438</td><td>Sales & Marketing</td><td>region_7</td><td>Master's & above</td><td>f</td><td>sourcing</td><td>1</td><td>35</td><td>5.0</td><td>8</td><td>1</td><td>0</td><td>49</td><td>0</td></tr><tr><td>1</td><td>65141</td><td>Operations</td><td>region_22</td><td>Bachelor's</td><td>m</td><td>other</td><td>1</td><td>30</td><td>5.0</td><td>4</td><td>0</td><td>0</td><td>60</td><td>0</td></tr><tr><td>2</td><td>7513</td><td>Sales & Marketing</td><td>region_19</td><td>Bachelor's</td><td>m</td><td>sourcing</td><td>1</td><td>34</td><td>3.0</td><td>7</td><td>0</td><td>0</td><td>50</td><td>0</td></tr><tr><td>3</td><td>2542</td><td>Sales & Marketing</td><td>region_23</td><td>Bachelor's</td><td>m</td><td>other</td><td>2</td><td>39</td><td>1.0</td><td>10</td><td>0</td><td>0</td><td>50</td><td>0</td></tr><tr><td>4</td><td>45845</td><td>Technology</td><td>region_25</td><td>Bachelor's</td><td>m</td><td>other</td><td>1</td><td>45</td><td>3.0</td><td>2</td><td>0</td><td>0</td><td>73</td><td>0</td></tr></tbody></table></div> | | employee_id | department | region | education | gender | recruitment_channel | no_of_trainings | age | previous_year_rating | length_of_service | Kpls_met >80% | awards_won? | avg_training_score | is_promoted | 0 | 65438 | Sales & Marketing | region_7 | Master's & above | f | sourcing | 1 | 35 | 5.0 | 8 | 1 | 0 | 49 | 0 | 1 | 65141 | Operations | region_22 | Bachelor's | m | other | 1 | 30 | 5.0 | 4 | 0 | 0 | 60 | 0 | 2 | 7513 | Sales & Marketing | region_19 | Bachelor's | m | sourcing | 1 | 34 | 3.0 | 7 | 0 | 0 | 50 | 0 | 3 | 2542 | Sales & Marketing | region_23 | Bachelor's | m | other | 2 | 39 | 1.0 | 10 | 0 | 0 | 50 | 0 | 4 | 45845 | Technology | region_25 | Bachelor's | m | other | 1 | 45 | 3.0 | 2 | 0 | 0 | 73 | 0 |
| | employee_id | department | region | education | gender | recruitment_channel | no_of_trainings | age | previous_year_rating | length_of_service | Kpls_met >80% | awards_won? | avg_training_score | is_promoted | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 65438 | Sales & Marketing | region_7 | Master's & above | f | sourcing | 1 | 35 | 5.0 | 8 | 1 | 0 | 49 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 65141 | Operations | region_22 | Bachelor's | m | other | 1 | 30 | 5.0 | 4 | 0 | 0 | 60 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | 7513 | Sales & Marketing | region_19 | Bachelor's | m | sourcing | 1 | 34 | 3.0 | 7 | 0 | 0 | 50 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | 2542 | Sales & Marketing | region_23 | Bachelor's | m | other | 2 | 39 | 1.0 | 10 | 0 | 0 | 50 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | 45845 | Technology | region_25 | Bachelor's | m | other | 1 | 45 | 3.0 | 2 | 0 | 0 | 73 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | |
|-----------------------|--|
| Handling Missing Data | <pre> # Replacing nan with mode print(df['education'].value_counts()) df['education']=df['education'].fillna(df['education'].mode()[0]) education Bachelor's 39078 Master's & above 14925 Below Secondary 805 Name: count, dtype: int64 [] # Replacing nan with mode print(df['previous_year_rating'].value_counts()) df['previous_year_rating']=df['previous_year_rating'].fillna(df['previous_year_rating'].mode()[0]) previous_year_rating 3.0 18618 5.0 11741 4.0 9877 1.0 6223 2.0 4225 Name: count, dtype: int64 </pre> |
| Data Transformation | <pre> # Feature mapping is done on education column from sklearn.preprocessing import LabelEncoder df['education'] = df['education'].replace(('Below Secondary', 'Bachelor', 'Master', ' & above'), ('0', '1', '2', '3')) lb = LabelEncoder() df['department'] = lb.fit_transform(df['department']) </pre> |
| Feature Engineering | Attached the codes in the final submission |
| Save Processed Data | - |