

# SAN FRANCISCO

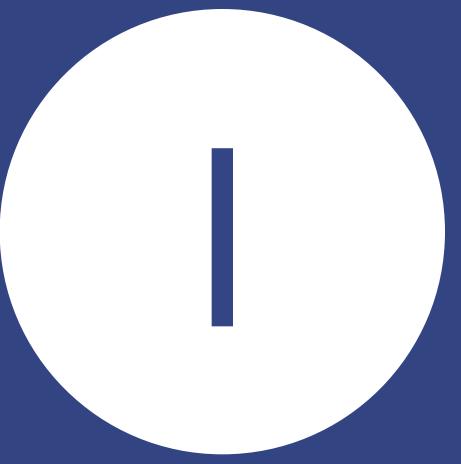
# WAGE GAP EDA

AMY ZHU



# OVERVIEW





# SITUATION & OBJECTIVE

# SITUATION



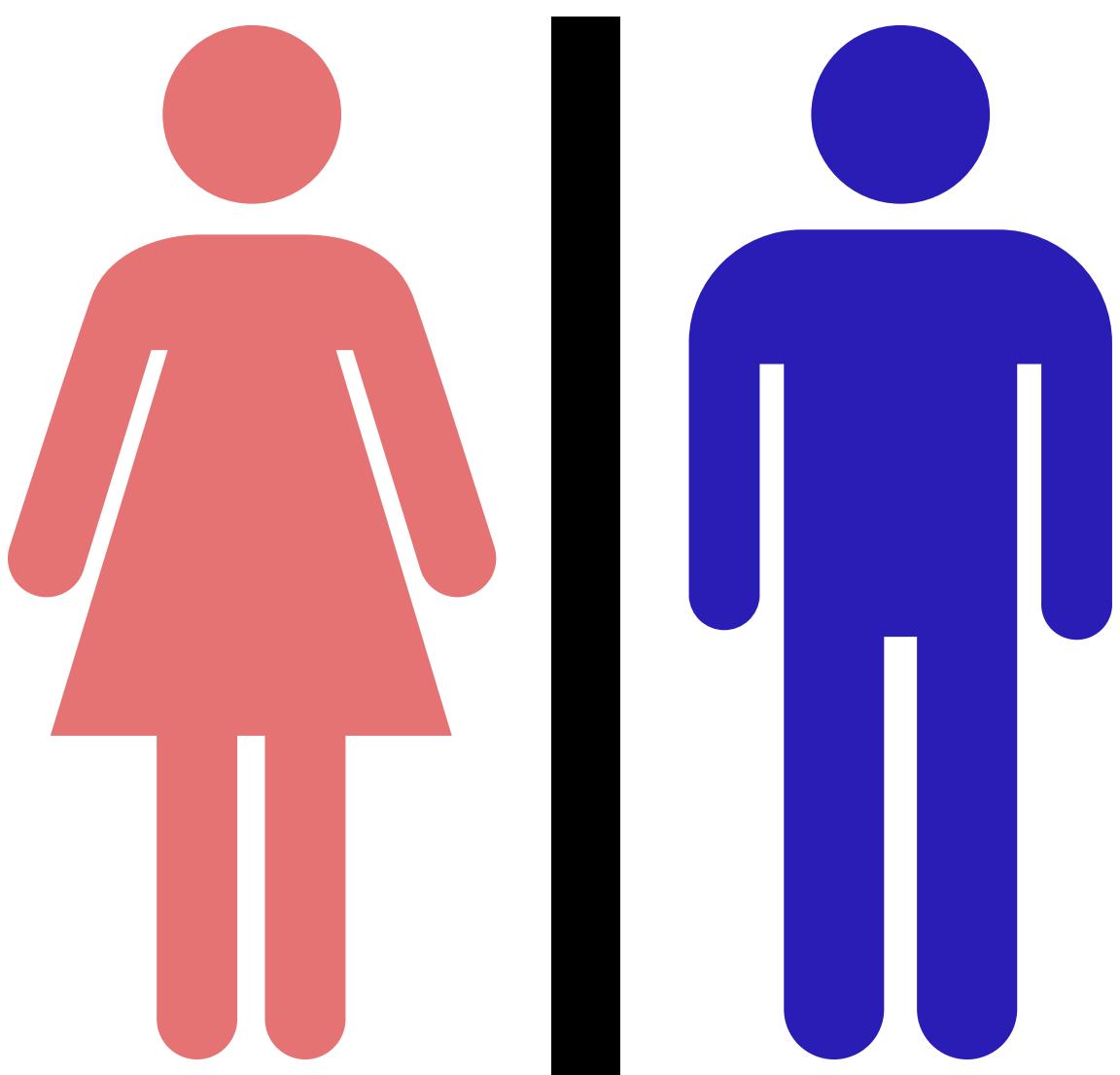
- Gavin is the **corrupt** CEO of a huge tech company, Hooli, located in Silicon Valley
- He is hiring employees to build Hooli's next big project, The Box 4
- Gavin wants to know if he could get away with hiring **female employees for less pay** than their equally skilled male counterparts

# OBJECTIVE

**Determine if there is a gender wage gap in the Bay Area**

Are women really paid less than men for the same job?

Are there differences in their salary breakdowns? (base & total)





# DATASET SYNOPSIS

# THE RAW DATA

This dataset was published by Kaggle with reports of salaries in the San Francisco Bay Area from 2011-2014

KEY VARIABLES  
Employee Name  
JobTitle  
BasePay  
Total Pay

| <b>Id</b> | <b>EmployeeName</b> | <b>JobTitle</b>                                | <b>BasePay</b> | <b>OvertimePay</b> | <b>OtherPay</b> | <b>Benefits</b> | <b>TotalPay</b> | <b>TotalPayBenefits</b> |
|-----------|---------------------|--|----------------|--------------------|-----------------|-----------------|-----------------|-------------------------|
| 1         | NATHANIEL FORD      | GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY | 167411.18      | 0.0                | 400184.25       | NaN             | 567595.43       | 567595.43               |
| 2         | GARY JIMENEZ        | CAPTAIN III (POLICE DEPARTMENT)                | 155966.02      | 245131.88          | 137811.38       | NaN             | 538909.28       | 538909.28               |
| 3         | ALBERT PARDINI      | CAPTAIN III (POLICE DEPARTMENT)                | 212739.13      | 106088.18          | 16452.6         | NaN             | 335279.91       | 335279.91               |
| 4         | CHRISTOPHER CHONG   | WIRE ROPE CABLE MAINTENANCE MECHANIC           | 77916.0        | 56120.71           | 198306.9        | NaN             | 332343.61       | 332343.61               |
| 5         | PATRICK GARDNER     | DEPUTY CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)  | 134401.6       | 9737.0             | 182234.59       | NaN             | 326373.19       | 326373.19               |
| 6         | DAVID SULLIVAN      | ASSISTANT DEPUTY CHIEF II                      | 118602.0       | 8601.0             | 189082.74       | NaN             | 316285.74       | 316285.74               |
| 7         | ALISON LEE          | BATTALION CHIEF /ETDE                          | 82402.01       | 80062.0            | 124426.14       | NaN             | 315081.05       | 315081.05               |

# DATA CLEANING

1. Drop irrelevant columns and rows with missing name values

2. Drop employee last name from dataset

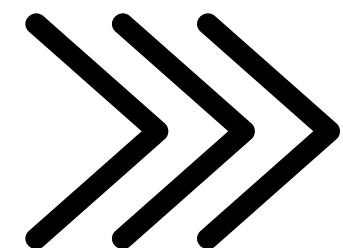
3. Drop jobs with <50 occurrences in dataset and drop salaries that are <0

# DETERMINING GENDER

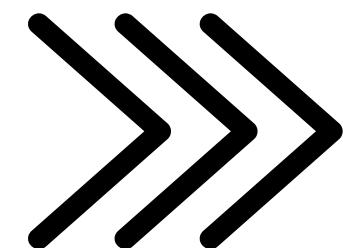
Use Python library **Gender Guesser** to assign gender to each employee based on their first name

## ASSIGNED GENDERS:

unknown  
mostly female  
mostly male  
androgynous  
**female**  
**male**



DROP ALL  
AMBIGUOUS  
GENDERS



COUNTS OF MALE  
VS. FEMALE  
EMPLOYEES

|        |       |
|--------|-------|
| male   | 69552 |
| female | 50236 |

# NEW TABLES (I)

|     |         | Name                               | JobTitle  | BasePay  | OvertimePay | OtherPay  | TotalPay  | TotalPayBenefits | Gender |
|-----|---------|------------------------------------|-----------|----------|-------------|-----------|-----------|------------------|--------|
| 22  | GEORGE  | CAPTAIN, FIRE SUPPRESSION          | 140546.88 | 93200.58 | 39955.25    | 273702.71 | 273702.71 |                  | male   |
| 24  | JOSEPH  | CAPTAIN, FIRE SUPPRESSION          | 140546.86 | 97868.77 | 31909.28    | 270324.91 | 270324.91 |                  | male   |
| 26  | JOHN    | CAPTAIN, FIRE SUPPRESSION          | 92080.80  | 40008.0  | 133695.76   | 265784.56 | 265784.56 |                  | male   |
| 33  | JOHN    | INSPECTOR III, (POLICE DEPARTMENT) | 104861.39 | 50227.61 | 103499.39   | 258588.39 | 258588.39 |                  | male   |
| 37  | JAMES   | INSPECTOR III, (POLICE DEPARTMENT) | 110661.20 | 31162.04 | 111446.2    | 253269.44 | 253269.44 |                  | male   |
| ... | ...     | ...                                | ...       | ...      | ...         | ...       | ...       | ...              | ...    |
| 563 | MAGALY  | FIREFIGHTER                        | 105934.68 | 58392.63 | 19409.94    | 183737.25 | 183737.25 |                  | female |
| 564 | ELISA   | NURSE MANAGER                      | 170596.03 | 0.0      | 13060.81    | 183656.84 | 183656.84 |                  | female |
| 566 | JACK    | FIREFIGHTER                        | 105934.66 | 59761.06 | 17944.28    | 183640.00 | 183640.00 |                  | male   |
| 568 | SUSAN   | SENIOR PHYSICIAN SPECIALIST        | 159624.81 | 0.0      | 23943.72    | 183568.53 | 183568.53 |                  | female |
| 569 | HERBERT | POLICE OFFICER III                 | 112401.11 | 11784.8  | 59346.05    | 183531.96 | 183531.96 |                  | male   |

MAIN TABLE WITH ADDED GENDER COLUMN

# NEW TABLES (2)

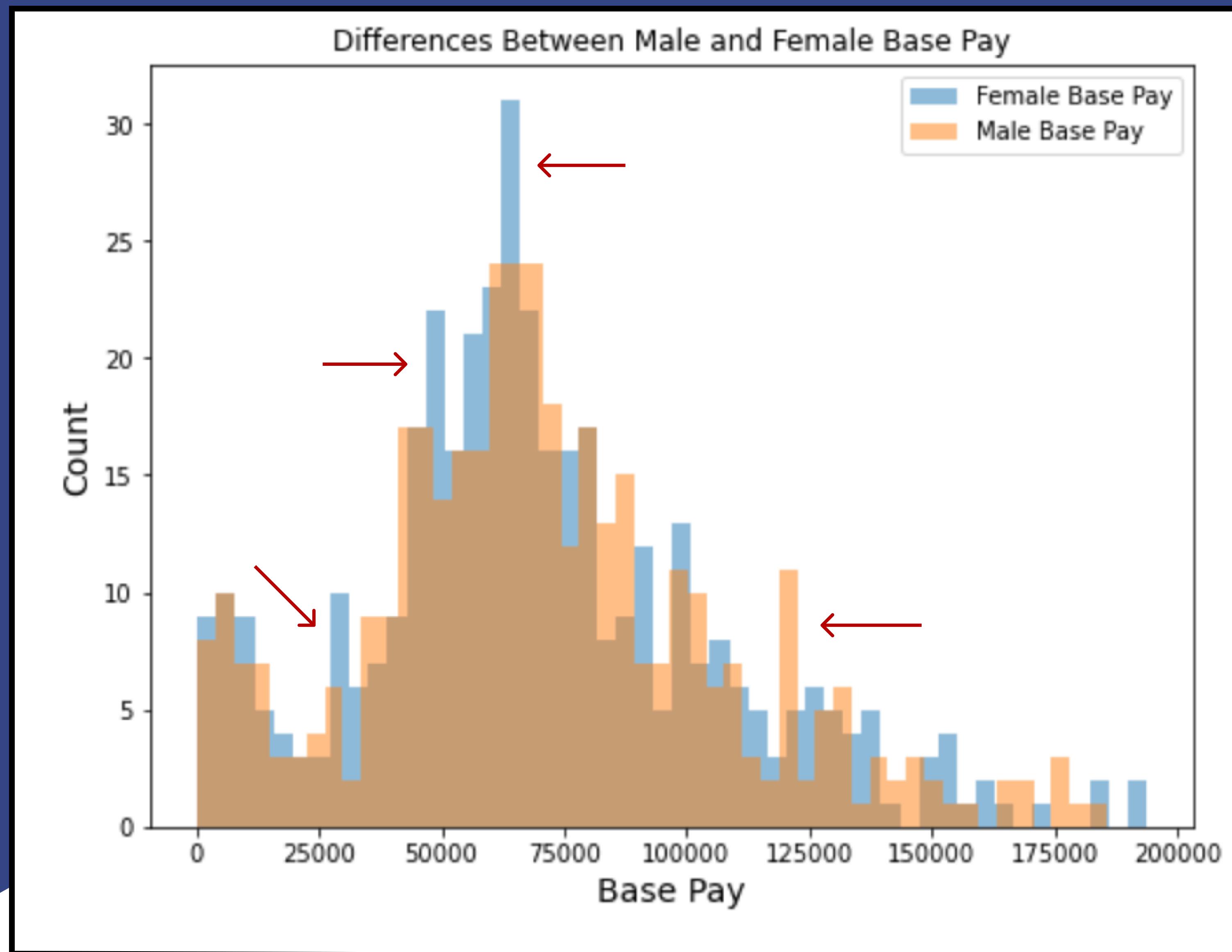
| JobTitle                     | FemaleBasePay | FemaleTotalPay | MaleBasePay  | MaleTotalPay | BasePayProp | TotalPayProp |
|------------------------------|---------------|----------------|--------------|--------------|-------------|--------------|
| ACCOUNT CLERK                | 42556.832391  | 43441.577391   | 46213.987333 | 46650.336667 | 92.086476   | 93.121680    |
| ADMINISTRATIVE ANALYST       | 64312.000513  | 64543.497949   | 64920.955294 | 65811.110000 | 99.062006   | 98.073863    |
| AIRPORT POLICE SERVICES AIDE | 49780.150345  | 57998.895000   | 52180.554167 | 61626.364688 | 95.399812   | 94.113770    |
| ASR Senior Office Specialist | 60577.726512  | 62563.650930   | 61041.301111 | 62685.674444 | 99.240556   | 99.805341    |
| ASSISTANT ENGINEER           | 80079.541111  | 80689.304815   | 83160.205227 | 84712.798864 | 96.295507   | 95.250430    |

NEW TABLE WITH AVERAGED PAY BASED ON GENDER AND  
JOB TITLE + PROPORTION OF FEMALE PAY TO MALE PAY

III

# VISUALIZATIONS

BASED ON COMBINED SALARY BY GENDER AND JOBLISTTABLE



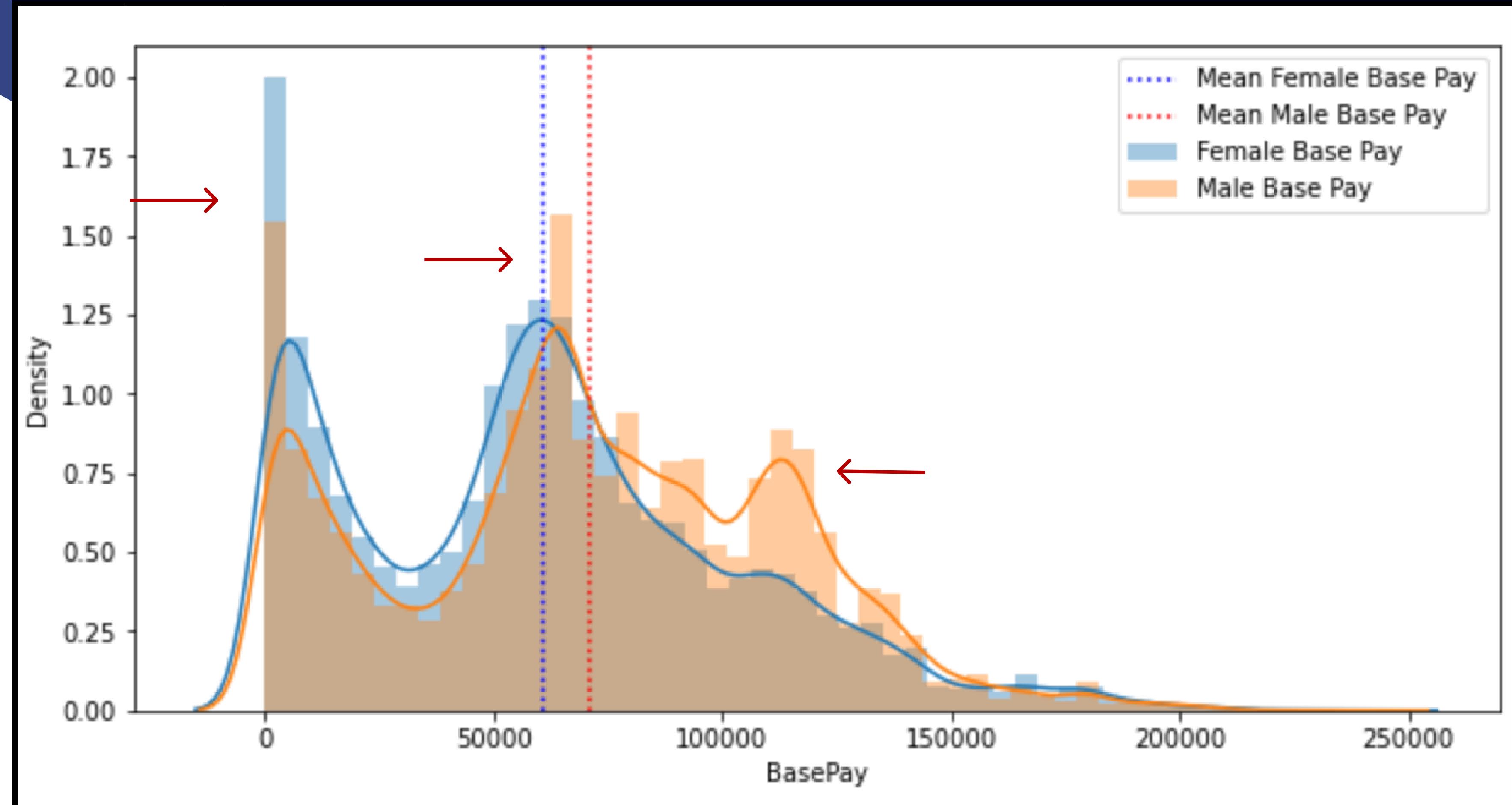
The frequency of female base salaries peak more at lower numbers than male salaries do

BASED ON INDIVIDUAL GENDER TABLES WITH CALCULATED AVG SALARY FOR ALL JOBS

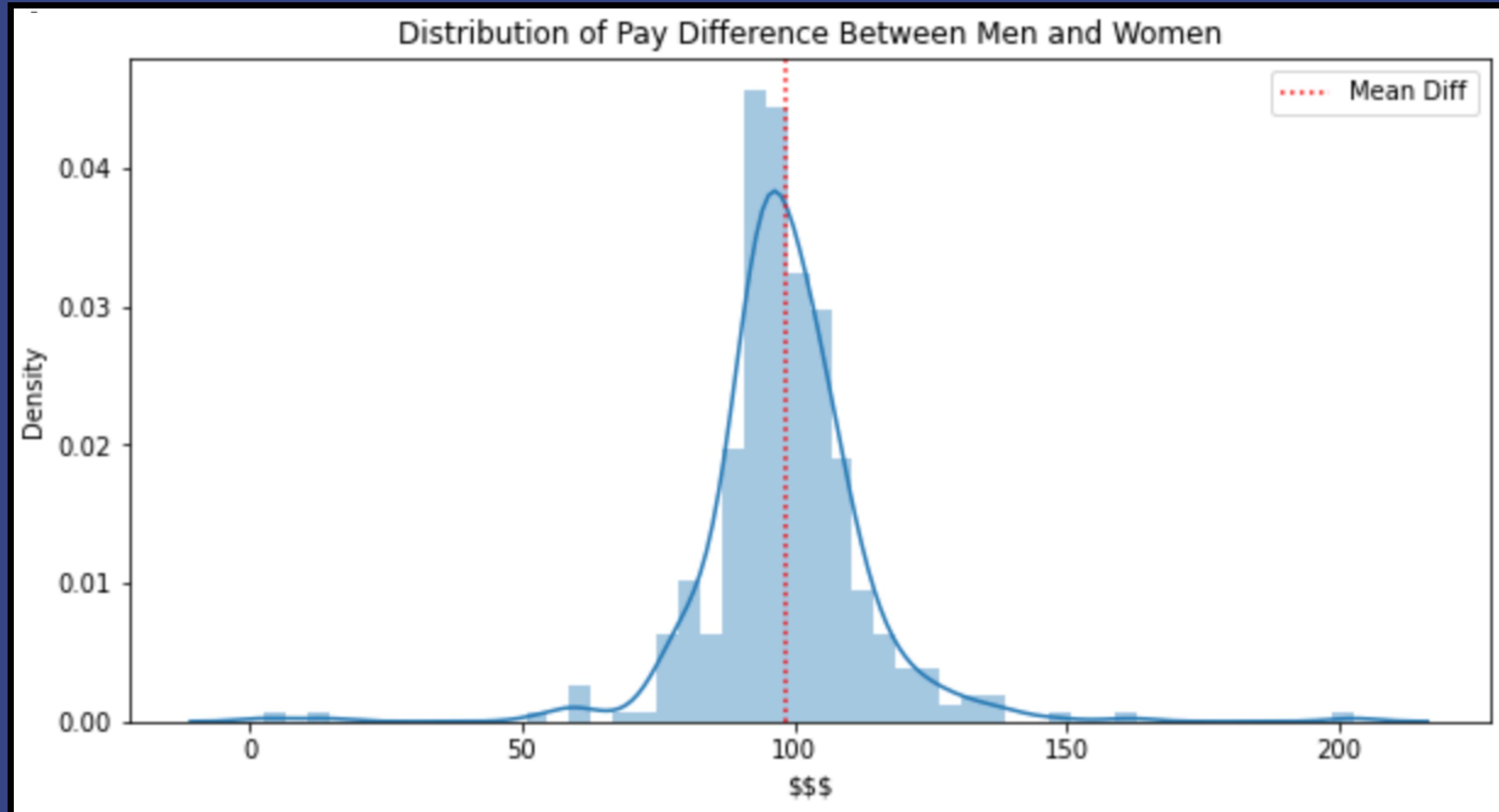
Men have higher mean base pay

Noticeable peak in male base pay past \$100,000

More women reported a base pay of \$0



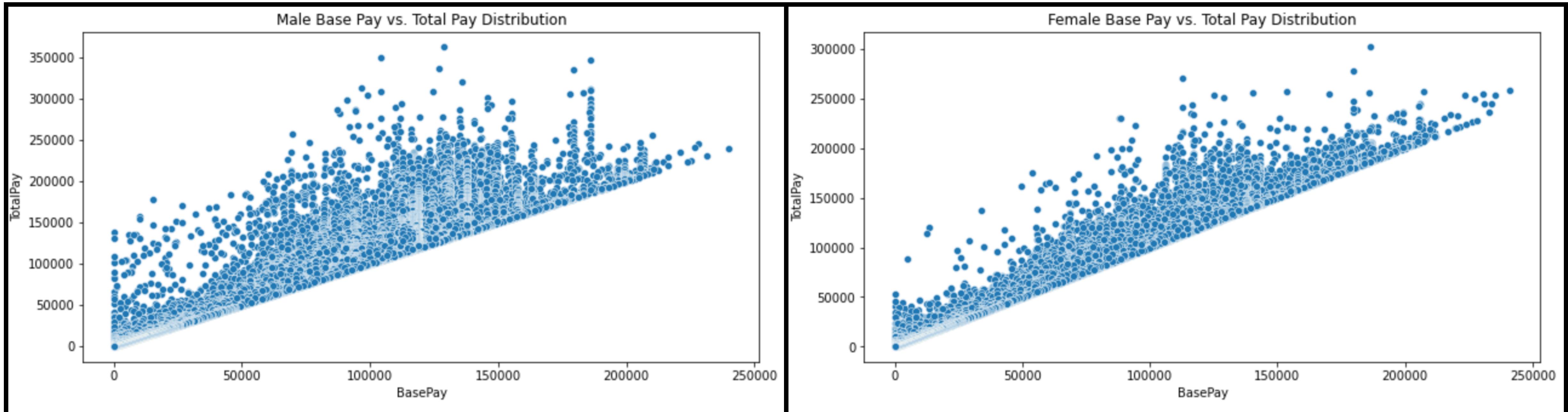
BASED ON TOTAL PAY PROPORTION VARIABLE



The mean proportion of female total pay to male total pay is ~98.37%

For every \$100 a man makes, a woman working the same job makes \$98.37

# BASE & TOTAL PAY



The spread of salaries is much wider for men, showing that **men usually earn more of their salary through bonuses and other types of pay than women**

- Could be related to the types of jobs men vs. women have



# ANALYSIS / INSIGHTS

# ARE WOMEN PAID LESS THAN MEN?

**Yes, but...**

- The difference in base pay is not entirely significant for the same job title
- Overall, this pay gap is likely due to the difference in the types of jobs they have  
(e.g. more women are social workers and more men are computer engineers)

# ARE THERE DIFFERENCES IN SALARY BREAKDOWNS?

- Men tend to earn more of their salary through **bonuses** than women do
- However, this could also be due to different jobs giving more bonuses than others



# CONCLUSION

# SITUATION VERDICT



Gavin will probably not be able to get away with significantly undercutting his female employees when comparing salaries of the same job title.

(At most, he can get away with paying female employees ~\$1.63 less per \$100 he pays his male employees without facing public backlash)

THANK YOU! ☺