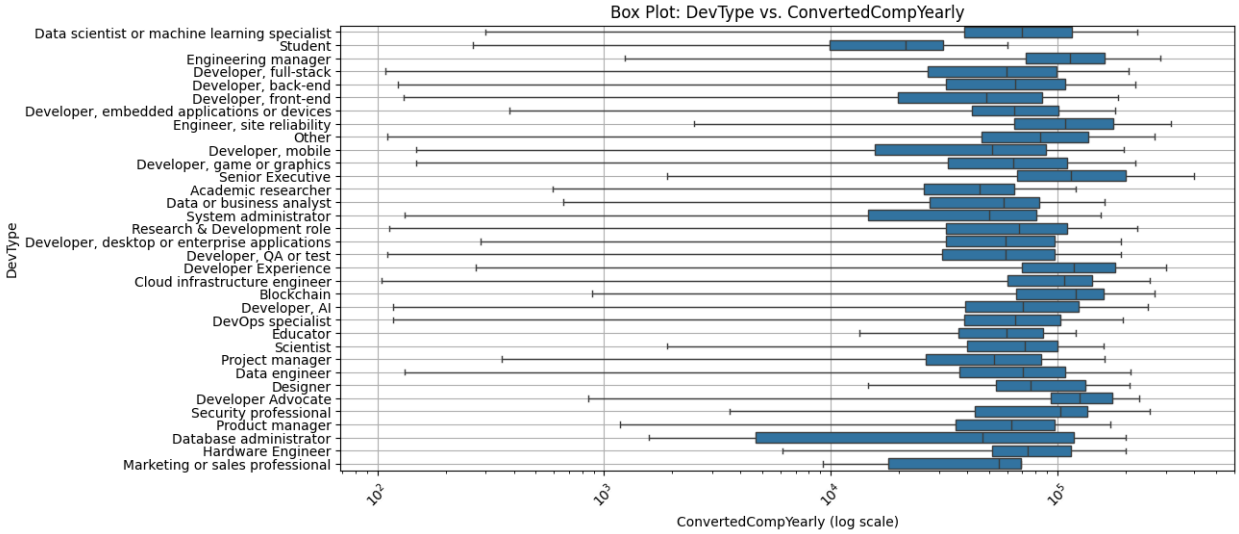
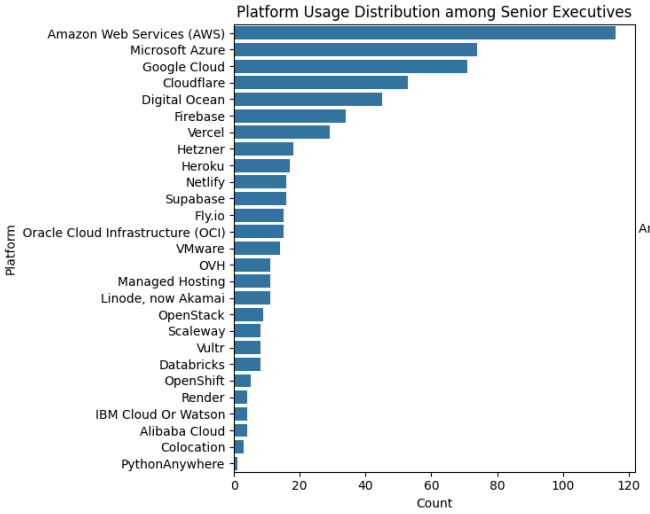
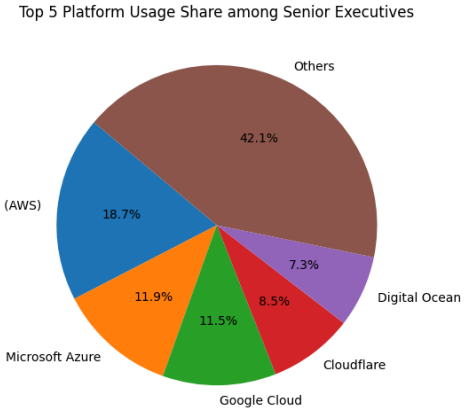
Q1 EDA

`DevType` v.s `ConvertedCompyYearly`



`PlatformHaveWorkedWith` by Senior Executive





Q2 Salary Differences Between Hybrid and Remote Job Modes

2(a)

Statistic	Hybrid	Remote
Count	5117.0	4758.0
Mean Salary	70739.73	79718.85
Std Dev	47701.05	56498.0
Min Salary	109.0	104.0
Max Salary	225861.0	226000.0

The IQR method was used to filter out extreme salary values beyond 1.5 times the interquartile range (below Q1 and above Q3).
Rationale: Salaries often have extreme outliers due to highly paid executives or very low salaries that may not reflect actual full-time earnings. Keeping only typical salary ranges improves the reliability of the comparison.

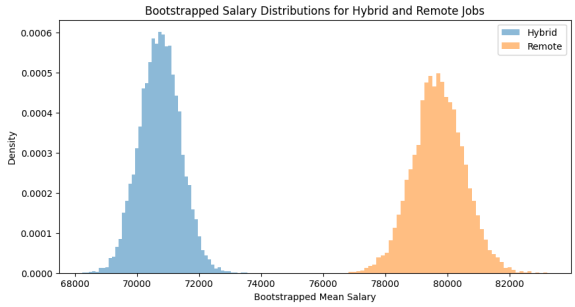
2(b). t-test
Assumptions:

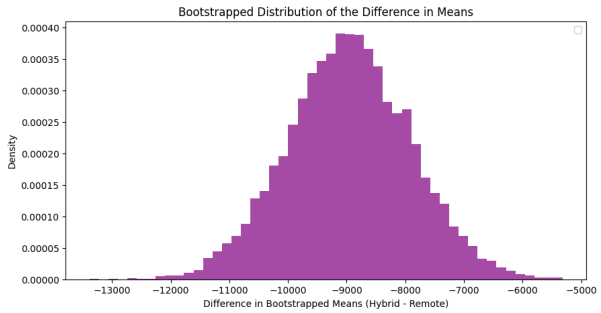
- Normal distribution: not satisfied, as shown visually in notebook; also by shapiro test, reject at 0.05 threshold
- Same var: not satisfied, by levene test, reject at 0.05 threshold

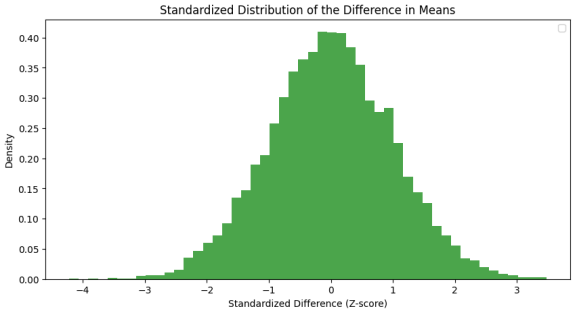
Results:

- t_manual: t = -8.55337
  t_builtin (equal var): t = -8.55337 p = 1.37315e-17
- Under not equal var assumption, t_manual = t_builtin

2(c). bootstrapped hybrid and remote salaries







2(d). t-test with bootstrapped data

- bootstrapped: t = -846.663 p = 0,
  Original data: t = -8.55337 p = 1.37315e-17
- Original data do not satisfied equal var and normal distribution assumptions and t-test performed on it (t = -8.55337 p = 1.37315e-17) is not reliable;

- Nevertheless, both p-value calculated from original and bootstrapped data are small, and suggest statistically significant difference in mean salaries of group remote and hybrid.

2(e). 95% CI with original v.s. bootstrapped data

- Formula-Based 95% CI for the difference (hybrid - remote) in means: [-11036.89693602 -6921.34463205]
- Bootstrap-Based 95% CI for the difference (hybrid - remote) in means: [-11024.86897858 -6879.95898405]
- the confidence intervals are quite close

2(f). median & Mood's test

- original p-value: 1.14501e-08
  bootstrapped p-value: 0
- Both p-value for Mood's test, calculated from original and bootstrapped data <0.05, and suggest statistically significant difference in median salaries of group remote and hybrid.

Q3 ANOVA

3(a)

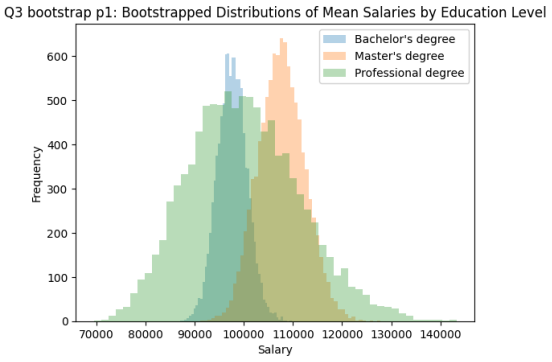
EdLevel	Count	Mean Salary	Std Dev	Min Salary
Associate degree	36	95137.03	43779.13	25443
Bachelor's degree	178	97633.73	40712.51	153
Master's degree	81	107661.68	42591.65	18000
Primary/elementary school	1	58154.0		58154
Professional degree	14	100151.36	43370.32	50885
Secondary school	12	81254.67	44833.08	13085
Some college/university study without earning a degree	49	92068.94	46043.2	19627
Something else	7	76836.57	29143.22	44343

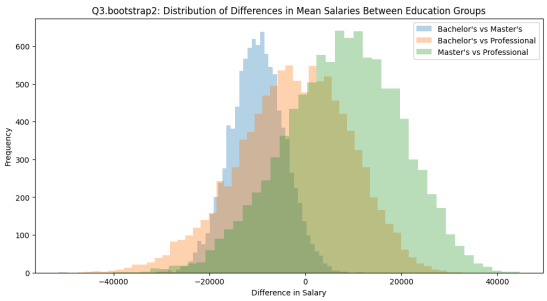
3(b). Assumptions

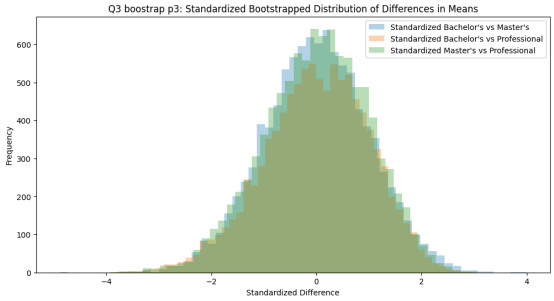
- Normality: not satisfied, see hist plot in notebook
- Equal variances: satisfied: (p-value>=0.05 for all 3 pair-wise levene tests)

p-value = 0.197096

3(c).







3(d)
bootstrapped ANOVA p-value: 0
ANOVA p-value: 0.197096

Since Bachelor’s degree, Master’s degree, Professional salaries are not normally distributed, ANOVA test result (p-value: 0.197096) on original data is not reliable. Bootstrapped data are normally distributed, and have small p-value which suggests statistically significant differences in mean salaries among the three education level groups.