12. Which parametet is highly correlated with salary?

```
In [1]: import pandas as pd
        import numpy as np
```

```
In [2]: dataset=pd.read_csv("Preplacementdata.csv")
        dataset
```

Out[2]:

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary	gender	ssc_b	hsc_b	hsc_s	degree_t	workex	special
0	1.0	67.00	91.00	58.00	55.0	58.80	270000.0	M	Others	Others	Commerce	Sci&Tech	No	
1	2.0	79.33	78.33	77.48	86.5	66.28	200000.0	M	Central	Others	Science	Sci&Tech	Yes	
2	3.0	65.00	68.00	64.00	75.0	57.80	250000.0	M	Central	Central	Arts	Comm&Mgmt	No	
3	4.0	56.00	52.00	52.00	66.0	59.43	265000.0	M	Central	Central	Science	Sci&Tech	No	
4	5.0	85.80	73.60	73.30	96.8	55.50	425000.0	M	Central	Central	Commerce	Comm&Mgmt	No	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
210	211.0	80.60	82.00	77.60	91.0	74.49	400000.0	M	Others	Others	Commerce	Comm&Mgmt	No	
211	212.0	58.00	60.00	72.00	74.0	53.62	275000.0	M	Others	Others	Science	Sci&Tech	No	
212	213.0	67.00	67.00	73.00	59.0	69.72	295000.0	M	Others	Others	Commerce	Comm&Mgmt	Yes	
213	214.0	74.00	66.00	58.00	70.0	60.23	204000.0	F	Others	Others	Commerce	Comm&Mgmt	No	
214	215.0	62.00	58.00	53.00	89.0	60.22	265000.0	M	Central	Others	Science	Comm&Mgmt	No	

215 rows × 15 columns
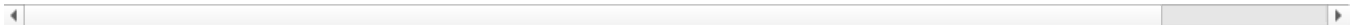
```
In [3]: dataset.isna().sum()
```

```
Out[3]: sl_no             0
        ssc_p             0
        hsc_p             0
        degree_p          0
        etest_p           0
        mba_p             0
        salary            0
        gender            0
        ssc_b             0
        hsc_b             0
        hsc_s             0
        degree_t          0
        workex            0
        specialisation    0
        status            0
        dtype: int64
```

```
In [4]: dir(dataset)
```

```
Out[4]: ['T',
         '_AXIS_LEN',
         '_AXIS_ORDERS',
         '_AXIS_TO_AXIS_NUMBER',
         '_HANDLED_TYPES',
         '__abs__',
         '__add__',
         '__and__',
         '__annotations__',
         '__array__',
         '__array_priority__',
         '__array_ufunc__',
         '__arrow_c_stream__',
         '__bool__',
         '__class__',
         '__contains__',
         '__copy__',
         '__dataframe__',
         '__dataframe_consortium_standard__',
         '__deepcopy__',
         '__delattr__',
         '__delitem__',
         '__dict__',
         '__dir__',
         '__divmod__',
```

```
'__doc__',
'__eq__',
'__finalize__',
'__floordiv__',
'__format__',
'__ge__',
'__getattr__',
'__getattribute__',
'__getitem__',
'__getstate__',
'__gt__',
'__hash__',
'__iadd__',
'__iand__',
'__ifloordiv__',
'__imod__',
'__imul__',
'__init__',
'__init_subclass__',
'__invert__',
'__ior__',
'__ipow__',
'__isub__',
'__iter__',
'__itruediv__',
'__ixor__',
'__le__',
'__len__',
'__lt__',
'__matmul__',
'__mod__',
'__module__',
'__mul__',
'__ne__',
'__neg__',
'__new__',
'__nonzero__',
'__or__',
'__pandas_priority__',
'__pos__',
'__pow__',
'__radd__',
'__rand__',
'__rdivmod__',
'__reduce__',
'__reduce_ex__',
'__repr__',
'__rfloordiv__',
'__rmatmul__',
'__rmod__',
'__rmul__',
'__ror__',
'__round__',
'__rpow__',
'__rsub__',
'__rtruediv__',
'__rxor__',
'__setattr__',
'__setitem__',
'__setstate__',
'__sizeof__',
'__str__',
'__sub__',
'__subclasshook__',
'__truediv__',
'__weakref__',
'__xor__',
'_accessors',
'_accum_func',
'_agg_examples_doc',
'_agg_see_also_doc',
'_align_for_op',
'_align_frame',
'_align_series',
'_append',
'_arith_method',
'_arith_method_with_reindex',
'_as_manager',
'_attrs',
'_box_col_values',
'_can_fast_transpose',
'_check_inplace_and_allows_duplicate_labels',
'_check_is_chained_assignment_possible',
```

```
'_check_label_or_level_ambiguity',
'_check_setitem_copy',
'_clear_item_cache',
'_clip_with_one_bound',
'_clip_with_scalar',
'_cmp_method',
'_combine_frame',
'_consolidate',
'_consolidate_inplace',
'_construct_axes_dict',
'_construct_result',
'_constructor',
'_constructor_from_mgr',
'_constructor_sliced',
'_constructor_sliced_from_mgr',
'_create_data_for_split_and_tight_to_dict',
'_data',
'_deprecate_downcast',
'_dir_additions',
'_dir_deletions',
'_dispatch_frame_op',
'_drop_axis',
'_drop_labels_or_levels',
'_ensure_valid_index',
'_find_valid_index',
'_flags',
'_flex_arith_method',
'_flex_cmp_method',
'_from_arrays',
'_from_mgr',
'_get_agg_axis',
'_get_axis',
'_get_axis_name',
'_get_axis_number',
'_get_axis_resolvers',
'_get_block_manager_axis',
'_get_bool_data',
'_get_cleaned_column_resolvers',
'_get_column_array',
'_get_index_resolvers',
'_get_item_cache',
'_get_label_or_level_values',
'_get_numeric_data',
'_get_value',
'_get_values_for_csv',
'_getitem_bool_array',
'_getitem_multilevel',
'_getitem_nocopy',
'_getitem_slice',
'_gotitem',
'_hidden_attrs',
'_indexed_same',
'_info_axis',
'_info_axis_name',
'_info_axis_number',
'_info_repr',
'_init_mgr',
'_inplace_method',
'_internal_names',
'_internal_names_set',
'_is_copy',
'_is_homogeneous_type',
'_is_label_or_level_reference',
'_is_label_reference',
'_is_level_reference',
'_is_mixed_type',
'_is_view',
'_is_view_after_cow_rules',
'_iset_item',
'_iset_item_mgr',
'_iset_not_inplace',
'_item_cache',
'_iter_column_arrays',
'_ixs',
'_logical_func',
'_logical_method',
'_maybe_align_series_as_frame',
'_maybe_cache_changed',
'_maybe_update_cacher',
'_metadata',
'_mgr',
'_min_count_stat_function',
'_needs_reindex_multi',
```

```
                    '_pad_or_backfill',
                    '_protect_consolidate',
                    '_reduce',
                    '_reduce_axis1',
                    '_reindex_axes',
                    '_reindex_multi',
                    '_reindex_with_indexers',
                    '_rename',
                    '_replace_columnwise',
                    '_repr_data_resource_',
                    '_repr_fits_horizontal_',
                    '_repr_fits_vertical_',
                    '_repr_html_',
                    '_repr_latex_',
                    '_reset_cache',
                    '_reset_cacher',
                    '_sanitize_column',
                    '_series',
                    '_set_axis',
                    '_set_axis_name',
                    '_set_axis_nocheck',
                    '_set_is_copy',
                    '_set_item',
                    '_set_item_frame_value',
                    '_set_item_mgr',
                    '_set_value',
                    '_setitem_array',
                    '_setitem_frame',
                    '_setitem_slice',
                    '_shift_with_freq',
                    '_should_reindex_frame_op',
                    '_slice',
                    '_stat_function',
                    '_stat_function_ddof',
                    '_take_with_is_copy',
                    '_to_dict_of_blocks',
                    '_to_latex_via_styler',
                    '_typ',
                    '_update_inplace',
                    '_validate_dtype',
                    '_values',
                    '_where',
                    'abs',
                    'add',
                    'add_prefix',
                    'add_suffix',
                    'agg',
                    'aggregate',
                    'align',
                    'all',
                    'any',
                    'apply',
                    'applymap',
                    'asfreq',
                    'asof',
                    'assign',
                    'astype',
                    'at',
                    'at_time',
                    'attrs',
                    'axes',
                    'backfill',
                    'between_time',
                    'bfill',
                    'bool',
                    'boxplot',
                    'clip',
                    'columns',
                    'combine',
                    'combine_first',
                    'compare',
                    'convert_dtypes',
                    'copy',
                    'corr',
                    'corrwith',
                    'count',
                    'cov',
                    'cummax',
                    'cummin',
                    'cumprod',
                    'cumsum',
                    'degree_p',
                    'degree_t',
```

```
                    'describe',
                    'diff',
                    'div',
                    'divide',
                    'dot',
                    'drop',
                    'drop_duplicates',
                    'droplevel',
                    'dropna',
                    'dtypes',
                    'duplicated',
                    'empty',
                    'eq',
                    'equals',
                    'etest_p',
                    'eval',
                    'ewm',
                    'expanding',
                    'explode',
                    'ffill',
                    'fillna',
                    'filter',
                    'first',
                    'first_valid_index',
                    'flags',
                    'floordiv',
                    'from_dict',
                    'from_records',
                    'ge',
                    'gender',
                    'get',
                    'groupby',
                    'gt',
                    'head',
                    'hist',
                    'hsc_b',
                    'hsc_p',
                    'hsc_s',
                    'iat',
                    'idxmax',
                    'idxmin',
                    'iloc',
                    'index',
                    'infer_objects',
                    'info',
                    'insert',
                    'interpolate',
                    'isetitem',
                    'isin',
                    'isna',
                    'isnull',
                    'items',
                    'iterrows',
                    'itertuples',
                    'join',
                    'keys',
                    'kurt',
                    'kurtosis',
                    'last',
                    'last_valid_index',
                    'le',
                    'loc',
                    'lt',
                    'map',
                    'mask',
                    'max',
                    'mba_p',
                    'mean',
                    'median',
                    'melt',
                    'memory_usage',
                    'merge',
                    'min',
                    'mod',
                    'mode',
                    'mul',
                    'multiply',
                    'ndim',
                    'ne',
                    'nlargest',
                    'notna',
                    'notnull',
                    'nsmallest',
```

```
'nunique',
'pad',
'pct_change',
'pipe',
'pivot',
'pivot_table',
'plot',
'pop',
'pow',
'prod',
'product',
'quantile',
'query',
'radd',
'rank',
'rdiv',
'reindex',
'reindex_like',
'rename',
'rename_axis',
'reorder_levels',
'replace',
'resample',
'reset_index',
'rfloordiv',
'rmod',
'rmul',
'rolling',
'round',
'rpow',
'rsub',
'rtruediv',
'salary',
'sample',
'select_dtypes',
'sem',
'set_axis',
'set_flags',
'set_index',
'shape',
'shift',
'size',
'skew',
'sl_no',
'sort_index',
'sort_values',
'specialisation',
'squeeze',
'ssc_b',
'ssc_p',
'stack',
'status',
'std',
'style',
'sub',
'subtract',
'sum',
'swapaxes',
'swaplevel',
'tail',
'take',
'to_clipboard',
'to_csv',
'to_dict',
'to_excel',
'to_feather',
'to_gbq',
'to_hdf',
'to_html',
'to_json',
'to_latex',
'to_markdown',
'to_numpy',
'to_orc',
'to_parquet',
'to_period',
'to_pickle',
'to_records',
'to_sql',
'to_stata',
'to_string',
'to_timestamp',
'to_xarray',
```

```
    'to_xml',
    'transform',
    'transpose',
    'truediv',
    'truncate',
    'tz_convert',
    'tz_localize',
    'unstack',
    'update',
    'value_counts',
    'values',
    'var',
    'where',
    'workex',
    'xs']
```

In [5]:
```python
from statsmodels.stats.outliers_influence import variance_inflation_factor
def calc_vif(X):
    vif=pd.DataFrame()
    vif["variables"]=X.columns
    vif["VIF"]=[variance_inflation_factor(X.values, i)for i in range(X.shape[1])]
    return(vif)
```

In [7]:
```python
calc_vif(dataset[['mba_p','salary']])
```

Out[7]:

	variables	VIF
0	mba_p	13.873041
1	salary	13.873041

In [8]:
```python
calc_vif(dataset[['ssc_p','salary']])
```

Out[8]:

	variables	VIF
0	ssc_p	11.624551
1	salary	11.624551

In [9]:
```python
calc_vif(dataset[['hsc_p','salary']])
```

Out[9]:

	variables	VIF
0	hsc_p	11.67638
1	salary	11.67638

In [10]:
```python
calc_vif(dataset[['degree_p','salary']])
```

Out[10]:

	variables	VIF
0	degree_p	12.545878
1	salary	12.545878

In [11]:
```python
calc_vif(dataset[['etest_p','salary']])
```

Out[11]:

	variables	VIF
0	etest_p	11.595795
1	salary	11.595795

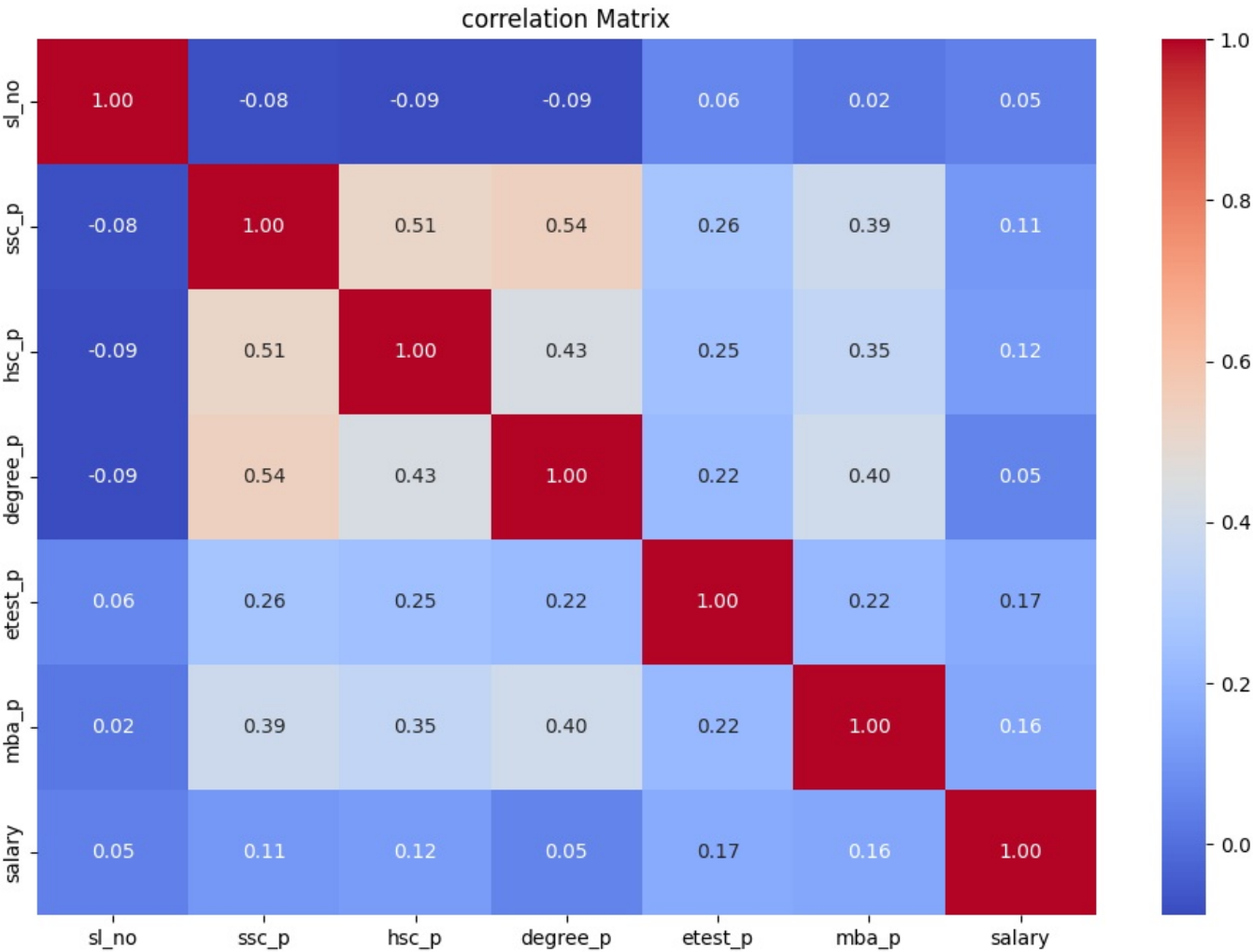all parameter are highly correlated with salary. But thr varaible mba_p and salary have a high correlation

13.plot any useful graph and explain it

In [16]:
```python
dataset.corr(numeric_only=True)
```

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
sl_no	1.000000	-0.078155	-0.085711	-0.088281	0.063636	0.022327	0.047270
ssc_p	-0.078155	1.000000	0.511472	0.538404	0.261993	0.388478	0.108669
hsc_p	-0.085711	0.511472	1.000000	0.434206	0.245113	0.354823	0.122921
degree_p	-0.088281	0.538404	0.434206	1.000000	0.224470	0.402364	0.053352
etest_p	0.063636	0.261993	0.245113	0.224470	1.000000	0.218055	0.169233
mba_p	0.022327	0.388478	0.354823	0.402364	0.218055	1.000000	0.155673
salary	0.047270	0.108669	0.122921	0.053352	0.169233	0.155673	1.000000

In [19]:
```python
import matplotlib.pyplot as plt
import seaborn as sns
correlation_matrix=dataset.corr(numeric_only=True)
plt.figure(figsize=(12,8))
sns.heatmap(correlation_matrix,annot=True,cmap='coolwarm',fmt='.2f')
plt.title ('correlation Matrix')
plt.show()
```


In [ ]:

In [ ]:

In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js