In [5]:

```
# -*- coding:utf-8 -*-
#对数据进行基本的探索
#返回缺失值个数以及最大最小值
import pandas as pd

datafile = 'air_data.csv'#航空公司原始数据,第一行是属性名

result = 'explore.xlsx'

data = pd. read_csv(datafile, encoding='utf-8')
#print(data.head())
explore = data.describe(percentiles = [], include = 'all').T
print(explore)
explore['null'] = len(data)-explore['count']

explore1 = explore[['null','max','min']]
explore1.columns = [u'空值数',u'最大值',u'最小值']#重命名列名

explore1.to_excel(result)
```

COUNT UNIFORM COUNT UNIFOR			:	4	C		-4.1
FPF_DATE	MEMBED NO			top	freq	mean	std \
FIRST_FLICHT_DATE	_						
GENDER 62985 2 P.B. 18134 NAN	——————————————————————————————————————						
PFP_TIER							
WORK_CITY 60719 3310 J*** 1750* NaN NaN WORK_COUNTRY 62962 118 CN 57748 NaN NaN AGE 62568 NaN NaN NaN 42,4763 9,8591 LOAD_TIME 62988 NaN NaN NaN NaN 11,0495 FLIGHT_COUNT 62988 NaN NaN NaN NaN 11,0495 BP_SLM 62988 NaN NaN NaN 10,025 16339.5 BP_SLM 62988 NaN NaN NaN NaN 10,00 0 EP_SLM_YR_1 62988 NaN NaN NaN NaN 266.99 1645.7 SLM_YR_1 62988 NaN NaN NaN 8109.46 38109.46 SLM_YR_1 62988 NaN NaN NaN 17123.9 20900.8 SIG_MSILW 62988 NaN NaN NaN NaN 17578.6 16339.5							
RORE, PROVINCE 59740 1185 J → 17507 NaN NaN MORE, COUNTRY 62962 118 CA → 7748 NaN NaN AGE 62568 NaN NaN NaN 42.4763 9.88591 LOAD, TIME 62988 NAN NAN NAN 11.8394 14.0495 BP_SUM 62988 NAN NAN NAN 10925.1 16339.5 BP_SUM YR_1 62988 NAN NAN NAN 266.69 1615.7 BP_SUM, YR_2 62988 NAN NAN NAN 526.38 8109.45 SUM_YR_1 62437 NAN NAN NAN 556.38 8109.45 SUM_YR_2 62880 NAN NAN NAN 1717.2 17578.6 SEC_M_SUM 62988 NAN NAN NAN 1717.2 17578.6 ASST_GLIGHT_OUNT 62988 NAN NAN NAN 120.14 123.1 195.2 AVG_B_SUM 62988							
WORK_COUNTRY 62962 118 CN 57748 NaN ANA AGE 62568 NaN NaN NAN 42.4763 9.88591 LOAD_TIME 62988 NaN NaN NaN NaN 11.6339.5 FLIGHT_COUNT 62988 NaN NaN NaN NaN 11.6339.5 EP_SUM_TR_1 62988 NaN NaN NaN NaN 1002.5 16339.5 EP_SUM_YR_1 62988 NaN NaN NaN NaN 263.69 1645.7 SIM_YR_1 62938 NaN NaN NaN NaN 555.38 8109.45 SIM_YR_2 62850 NaN NaN NaN NaN 17123.9 20960.8 SEC,MSUM 62988 NaN NaN NaN NaN 17277.2 17578.6 AST_FLIGHT_DATE 62988 NaN NaN NaN NaN 17578.6 AST_GERM 62988 NaN NaN NaN <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>							
AGE COAD_TIME 62568 Nan Nan Nan 42.4763 9.88591 LOAD_TIME 62988 1 2014/03/31 62988 Nan Nan 11.8394 1.0495 BP_SUM 62988 Nan Nan Nan 10925.1 16339.5 EP_SUM_YR_1 62988 Nan Nan Nan 10925.1 16339.5 EP_SUM_YR_2 62988 Nan Nan Nan 525.69 1615.7 SUM_YR_1 62437 Nan Nan Nan 5601.03 8703.36 SEC_KM_SUM 62988 Nan Nan Nan 17172.2 20960.8 SEC_KM_SUM 62988 Nan Nan Nan 11712.3 20960.8 SEC_KM_SUM 62988 Nan Nan Nan 1277.7 17578.6 ASS_CE_KM_SUM 62988 Nan Nan Nan 12014/03/31 599 Nan AVG_STURING 62988 Nan Nan Nan	_						
LOAD_TIME	_						
FLIGHT_COUNT							
BP_SIM 62988 NAN NAN NAN 10925.1 16339.5 EP_SIM_YR_1 62988 NAN NAN NAN A0 0 EP_SIM_YR_1 62988 NAN NAN NAN 26.66 9 164.57 SUM_YR_1 62987 NAN NAN NAN 5355.38 8109.45 SUM_YR_1 62988 NAN NAN NAN 17123.9 20960.8 SUE,GM_SUM 62988 NAN NAN NAN 12777.2 17578.6 LAST_FLIGHT_DATE 62988 NAN NAN NAN NAN 1.7577.6 17578.6 LAST_FLIGHT_COUNT 62988 NAN NAN NAN 1.7517.5 1.7578.6 LAST_TO_END 62988 NAN NAN NAN 1.21.145 15.9.573 LAST_TO_END 62988 NAN NAN NAN 1.12.145 15.9.573 LAST_TO_END 62988 NAN NAN NAN 166.034 17.3192 <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>							
EP_SLM_YR.1 62988 NAN NAN NAN 0.0 EP_SUM_YR.2 62988 NAN NAN NAN 265. 69 1645. 7 SUM_YR.2 62850 NAN NAN NAN 535. 38 8109. 45 SUM_YR.2 62850 NAN NAN NAN 5604. 03 8703. 36 SUM_YR.2 62988 NAN NAN NAN NAN 1777. 2 17578. 6 LAST_FLIGHT_DATE 62988 NAN NAN NAN 154215 1.787 AVG_FLIGHT_COUNT 62988 NAN NAN NAN 1421.44 2083. 12 BEGIS_TO_FIRST 62988 NAN NAN NAN 176. 12 183.822 MCG_TITERVAL 62988 NAN NAN NAN 176. 12 183.822 MAY_INTERVAL 62988 NAN NAN NAN 166.034 123.397 ADD_POINTS_SUM_YR_1 62988 NAN NAN NAN 504.317 3956.08	_						
EP_SIM_YN_2 62988 NAN NAN NAN 265. 69 1645. 7 SIM_YR_1 62437 NAN NAN NAN 535.38 8109.45 SIM_YR_2 62850 NAN NAN NAN 5604.03 8703.36 SEC_MLSUM 62988 NAN NAN NAN 17123.9 20960.8 WEIGHTED_SEC_KM 62988 NAN NAN NAN 17123.9 20960.8 MELGHTED_SEC_KM 62988 NAN NAN NAN 17123.9 20960.8 MELGHTED_SEC_KM 62988 NAN NAN NAN NAN 17173.7 17578.6 AVE_DES_CM 62988 NAN NAN NAN NAN 1421.44 2038.12 LAST_O_FIRST 62988 NAN NAN NAN NAN 176.12 183.822 AVE_INTERVAL 62988 NAN NAN NAN 176.149 175.179 AND_POINTS_SUM_YR_1 62988 NAN NAN NAN							
SUM_YR_1							
SUN_YR_2							
SEC_KM_SUM	_						
WEIGHTED_SEG_KM							
LAST_FILIGHT_DATE							
AVG_FLIGHT_COUNT							
AVG_BP_SUM							
BEGIN_TO_FIRST 62988 NaN NaN NaN 120.145 159.573 LAST_TO_END 62988 NaN NaN NaN 176.12 138.822 AVG_INTERVAL 62988 NaN NaN NaN 67.7498 77.5179 MAX_INTERVAL 62988 NaN NaN NaN 1660.034 123.397 ADD_POINTS_SUM_YR_1 62988 NaN NaN NaN 540.317 3956.08 ADD_POINTS_SUM_YR_2 62988 NaN NaN NaN 0.319775 1.136 avg_discount 62988 NaN NaN NaN 0.319775 1.136 avg_discount 62988 NaN NaN NaN 0.319775 1.136 avg_discount 62988 NaN NaN NaN 0.721558 0.185427 PIY_Flight_Count 62988 NaN NaN NaN 5.76626 7.21092 LIY_Flight_Count 62988 NaN NaN NaN 5366.72 8537.77 LIY_BP_SUM 62988 NaN NaN NaN 1355.01 7868.48 EI_Add_Point_Sum 62988 NaN NaN NaN 1355.01 7868.48 EI_Add_Point_Sum 62988 NaN NaN NaN 1620.7 8294.4 LIY_EL_Add_Points 62988 NaN NaN NaN 1620.7 8294.4 LIY_Points_Sum 62988 NaN NaN NaN 1080.38 5639.86 Point_Sum 62988 NaN NaN NaN 0.486419 0.319105 Ration_PIY_Flight_Count 62988 NaN NaN NaN 0.513581 0.319105 Ration_PIY_Flight_Count 62988 NaN NaN NaN 0.513581 0.319105 Ration_PIY_Flight_Count 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN NaN 0.468422 0.338956 Point_NotFlight NaN NaN NaN NaN NaN NaN 0.468422 0.338956 Point_NotFlight NaN							
LAST_TO_END	_						
AVG_INTERVAL 62988 NaN NaN NaN 166. 034 123. 397							
MAX_INTERVAL 62988 NaN NaN NaN 166.034 123.397 ADD_POINTS_SUM_YR_1 62988 NaN NaN NaN 540.317 3956.08 ADD_POINTS_SUM_YR_2 62988 NaN NaN NaN NaN 540.317 3956.08 ADD_POINTS_SUM_YR_2 62988 NaN NaN NaN NaN 0.319775 1.136 avg_discount 62988 NaN NaN NaN 0.721558 0.185427 PIY_Flight_Count 62988 NaN NaN NaN 0.721558 0.185427 PIY_Flight_Count 62988 NaN NaN NaN 5.76626 7.21092 LIY_Flight_Count 62988 NaN NaN NaN 5.76626 7.21092 LIY_Flight_Count 62988 NaN NaN NaN 5366.72 8537.77 LIY_BP_SUM 62988 NaN NaN NaN 5558.36 9351.96 EP_SUM 62988 NaN NaN NaN 5558.36 9351.96 EP_SUM 62988 NaN NaN NaN 265.69 1645.7 ADD_Point_SUM 62988 NaN NaN NaN 1355.01 7868.48 Eli_Add_Point_Sum 62988 NaN NaN NaN 1620.7 78294.4 LIY_ELi_Add_Points 62988 NaN NaN NaN 1620.7 78294.4 LIY_ELi_Add_Points 62988 NaN NaN NaN 1620.7 78294.4 LIY_Eli_JAdd_Points 62988 NaN NaN NaN 1633.38 5639.86 Points_Sum 62988 NaN NaN NaN 1633.31 12601.8 Ration_LIY_Flight_Count 62988 NaN NaN NaN 0.486419 0.319105 Ration_LIY_Flight_Count 62988 NaN NaN NaN 0.513581 0.319105 Ration_LIY_BPS 62988 NaN NaN NaN 0.522293 0.339632 Ration_LIY_BPS 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN NaN 0.468422 0.338956 Point_NotFlight NaN							
ADD_POINTS_SUM_YR_1 62988	——————————————————————————————————————						
ADD_POINTS_SUM_YR_2	——————————————————————————————————————						
EXCHANGE_COUNT							
avg_discount 62988 NaN NaN NaN 0. 721558 0. 185427 PIY_Flight_Count 62988 NaN NaN NaN 5. 76626 7. 21092 LIY_Flight_Count 62988 NaN NaN NaN 6.07316 8. 17513 PIY_BP_SUM 62988 NaN NaN NaN 5366. 72 8537. 77 LIY_BP_SUM 62988 NaN NaN NaN 5366. 72 8537. 77 LIY_BP_SUM 62988 NaN NaN NaN 5366. 72 8537. 77 LIY_B_SUM 62988 NaN NaN NaN NaN 536.69 9351. 96 EP_SUM 62988 NaN NaN NaN NaN 1355. 01 7868. 48 EI_Add_Points 62988 NaN NaN NaN 1620. 7 8294. 4 LIY_Points_Sum 62988 NaN NaN NaN 1620. 38 5639. 86 Points_Sum 62988 NaN NaN NaN NaN<							
P11_Flight_Count 62988 NaN NaN NaN 5. 76626 7. 21092 L1Y_Flight_Count 62988 NaN NaN NaN 6.07316 8. 17513 P1Y_BP_SUM 62988 NaN NaN NaN 5366. 72 8537. 77 L1Y_BP_SUM 62988 NaN NaN NaN 5558. 36 9351. 96 EP_SUM 62988 NaN NaN NaN NaN 265. 69 1645. 7 ADD_Point_SUM 62988 NaN NaN NaN NaN 1355. 01 7868. 48 E1i_Add_Point_Sum 62988 NaN NaN NaN 1620. 7 8294. 4 L1Y_ELi_Add_Points 62988 NaN NaN NaN 1620. 7 8294. 4 L1Y_ELi_Add_Points 62988 NaN NaN NaN 1080. 38 5639. 86 Points_Sum 62988 NaN NaN NaN 12545. 8 20507. 8 L1Y_Points_Sum 62988 NaN NaN NaN <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>							
L1Y_Flight_Count 62988 NaN NaN NaN 5366.72 8537.77 L1Y_BP_SUM 62988 NaN NaN NaN 5366.72 8537.77 L1Y_BP_SUM 62988 NaN NaN NaN 5558.36 9351.96 EP_SUM 62988 NaN NaN NaN 5558.36 9351.96 EP_SUM 62988 NaN NaN NaN 265.69 1645.7 ADD_Point_SUM 62988 NaN NaN NaN 1355.01 7868.48 Eli_Add_Point_Sum 62988 NaN NaN NaN 1620.7 8294.4 L1Y_ELi_Add_Points 62988 NaN NaN NaN 1620.7 8294.4 L1Y_ELi_Add_Points 62988 NaN NaN NaN 1080.38 5639.86 Points_Sum 62988 NaN NaN NaN 1080.38 5639.86 Points_Sum 62988 NaN NaN NaN 1080.38 5639.86 Ration_L1Y_Flight_Count 62988 NaN NaN NaN 0.486419 0.319105 Ration_P1Y_Flight_Count 62988 NaN NaN NaN 0.486419 0.319105 Ration_P1Y_BPS 62988 NaN NaN NaN 0.513581 0.319105 Ration_P1Y_BPS 62988 NaN NaN NaN 0.522293 0.339632 Ration_L1Y_BPS 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN NaN 0.468422 0.338956 Point_NotFlight NaN NaN NaN NaN NaN NaN NaN NaN NaN Na	- -						
P1Y_BP_SUM 62988 NaN NaN NaN 5366.72 8537.77 L1Y_BP_SUM 62988 NaN NaN NaN 5558.36 9351.96 EP_SUM 62988 NaN NaN NaN 265.69 1645.7 ADD_Point_SUM 62988 NaN NaN NaN 1355.01 7868.48 Eli_Add_Point_Sum 62988 NaN NaN NaN 1620.7 78294.4 L1Y_Eli_Add_Points 62988 NaN NaN NaN 1080.38 5639.86 Points_Sum 62988 NaN NaN NaN 1080.38 5639.86 Points_Sum 62988 NaN NaN NaN 12545.8 20507.8 L1Y_Points_Sum 62988 NaN NaN NaN 12601.8 8 Ration_L1Y_Flight_Count 62988 NaN NaN NaN 0.319105 Ration_L1Y_BPS 62988 NaN NaN NaN NaN 0.319105 Ration_L1Y_							
L1Y_BP_SUM 62988 NaN NaN NaN 5558.36 9351.96 EP_SUM 62988 NaN NaN NaN 265.69 1645.7 ADD_Point_SUM 62988 NaN NaN NaN 1355.01 7868.48 Eli_Add_Point_Sum 62988 NaN NaN NaN 1620.7 8294.4 L1Y_ELi_Add_Points 62988 NaN NaN NaN 1080.38 5639.86 Points_Sum 62988 NaN NaN NaN 12545.8 20507.8 L1Y_Points_Sum 62988 NaN NaN NaN 12545.8 20507.8 L1Y_Points_Sum 62988 NaN NaN NaN NaN 6638.74 12601.8 Ration_L1Y_Flight_Count 62988 NaN NaN NaN NaN 0.486419 0.319105 Ration_P1Y_Flight_Count 62988 NaN NaN NaN NaN 0.513581 0.319105 Ration_P1Y_BPS 62988 NaN NaN NaN NaN 0.522293 0.339632 Ration_L1Y_BPS 62988 NaN NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN NaN 0.468422 0.338956 Point_TotFlight_DATE NaN NaN NaN NaN NaN NaN NaN NaN NaN Na							
EP_SUM 62988 NaN NaN NaN 265.69 1645.7 ADD_Point_SUM 62988 NaN NaN NaN 1355.01 7868.48 Eli_Add_Point_Sum 62988 NaN NaN NaN 1620.7 8294.4 L1Y_Eli_Add_Points 62988 NaN NaN NaN 1080.38 5639.86 Points_Sum 62988 NaN NaN NaN 12545.8 20507.8 L1Y_Points_Sum 62988 NaN NaN NaN 12601.8 8 Ration_L1Y_Flight_Count 62988 NaN NaN NaN 0.319105 8 Ration_P1Y_Flight_Count 62988 NaN NaN NaN 0.513581 0.319105 Ration_P1Y_BPS 62988 NaN NaN NaN 0.522293 0.339632 Ration_L1Y_BPS 62988 NaN NaN NaN NaN 0.346422 0.338956 Point_NotFlight 62988 NaN NaN NaN NaN							
ADD_Point_SUM 62988 NaN NaN NaN 1355.01 7868.48 Eli_Add_Point_Sum 62988 NaN NaN NaN 1620.7 8294.4 L1Y_ELi_Add_Points 62988 NaN NaN NaN 1080.38 5639.86 Points_Sum 62988 NaN NaN NaN 12545.8 20507.8 L1Y_Points_Sum 62988 NaN NaN NaN 12545.8 20507.8 L1Y_Points_Sum 62988 NaN NaN NaN 0.486419 0.319105 Ration_l1Y_Flight_Count 62988 NaN NaN NaN 0.513581 0.319105 Ration_P1Y_Flight_Count 62988 NaN NaN NaN 0.512293 0.339632 Ration_P1Y_BPS 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight NaN NaN NaN NaN NaN 0.468422 MEMBER_NO 1 31494.5 62988 FFP_DATE NaN NaN NaN NaN NaN FIRST_FLIGHT_DATE NaN NaN NaN NaN FIRST_FLIGHT_DATE NaN NaN NaN NaN NaN FFP_TIER 4 4 6 WORK_CITY NaN NaN NaN NaN WORK_CITY NaN NaN NaN NaN WORK_COUNTRY NaN NaN NaN AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0							
Eli_Add_Point_Sum 62988 NaN NaN NaN 1620.7 8294.4 L1Y_ELi_Add_Points 62988 NaN NaN NaN 1080.38 5639.86 Points_Sum 62988 NaN NaN NaN 12545.8 20507.8 L1Y_Points_Sum 62988 NaN NaN NaN 12545.8 20507.8 L1Y_Points_Sum 62988 NaN NaN NaN NaN 12545.8 20507.8 L1Y_Points_Sum 62988 NaN NaN NaN NaN 6638.74 12601.8 Ration_L1Y_Flight_Count 62988 NaN NaN NaN 0.486419 0.319105 Ration_P1Y_BPS 62988 NaN NaN NaN NaN 0.319105 Ration_L1Y_BPS 62988 NaN NaN NaN NaN 0.33956 Point_NotFlight 62988 NaN NaN <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>							
L1Y_ELi_Add_Points 62988 NaN NaN NaN 1080.38 5639.86 Points_Sum 62988 NaN NaN NaN 12545.8 20507.8 L1Y_Points_Sum 62988 NaN NaN NaN 6638.74 12601.8 Ration_L1Y_Flight_Count 62988 NaN NaN NaN 0.486419 0.319105 Ration_P1Y_Flight_Count 62988 NaN NaN NaN 0.513581 0.319105 Ration_P1Y_Flight_Count 62988 NaN NaN NaN 0.513581 0.319105 Ration_P1Y_BPS 62988 NaN NaN NaN 0.522293 0.339632 Ration_L1Y_BPS 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight NaN NaN NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN NaN NaN NaN NaN NaN NaN N							
Points_Sum 62988 NaN NaN NaN 12545.8 20507.8 L1Y_Points_Sum 62988 NaN NaN NaN 6638.74 12601.8 Ration_L1Y_Flight_Count 62988 NaN NaN 0.486419 0.319105 Ration_P1Y_Flight_Count 62988 NaN NaN 0.513581 0.319105 Ration_P1Y_BPS 62988 NaN NaN 0.522293 0.339632 Ration_L1Y_BPS 62988 NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN NaN NaN MEMBER_NO 1 31494.5 62988 8 4 4 6 FFP_DATE <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>							
L1Y_Points_Sum 62988 NaN NaN NaN 6638.74 12601.8 Ration_L1Y_Flight_Count 62988 NaN NaN NaN 0.486419 0.319105 Ration_P1Y_Flight_Count 62988 NaN NaN NaN 0.513581 0.319105 Ration_P1Y_BPS 62988 NaN NaN NaN 0.522293 0.339632 Ration_L1Y_BPS 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 0.319105 0.339632 Ration_L1Y_BPS 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN NaN NaN FIRST_FLIGHT_DATE NaN NaN NaN NaN NaN<							
Ration_L1Y_Flight_Count 62988 NaN NaN NaN 0.486419 0.319105 Ration_P1Y_Flight_Count 62988 NaN NaN NaN 0.513581 0.319105 Ration_P1Y_BPS 62988 NaN NaN NaN 0.522293 0.339632 Ration_L1Y_BPS 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 2.72815 7.36416 **MEMBER_NO** **Interpretable **Inter	_						
Ration_P1Y_F1ight_Count 62988 NaN NaN NaN 0.513581 0.319105 Ration_P1Y_BPS 62988 NaN NaN NaN 0.468422 0.338956 Ration_L1Y_BPS 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 2.72815 7.36416 MEMBER_NotA 1 31494.5 62988 62988 8							
Ration_P1Y_BPS 62988 NaN NaN NaN 0.322293 0.339632 Ration_L1Y_BPS 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 2.72815 7.36416 MEMBER_NO 1 31494.5 62988 7288 7288 7288 7288 7288							
Ration_L1Y_BPS 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 0.468422 0.338956 Point_NotFlight 62988 NaN NaN NaN 2.72815 7.36416 MEMBER_NO 1 31494.5 62988							
Point_NotFlight 62988 NaN NaN NaN 2.72815 7.36416 MEMBER_NO 1 31494.5 62988 6298 62988 6298 6298 6298 <							
min 50% max MEMBER_NO 1 31494. 5 62988 FFP_DATE NaN NaN NaN FIRST_FLIGHT_DATE NaN NaN NaN GENDER NaN NaN NaN FFP_TIER 4 4 6 WORK_CITY NaN NaN NaN WORK_PROVINCE NaN NaN NaN WORK_COUNTRY NaN NaN NaN AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0							
MEMBER_NO 1 31494. 5 62988 FFP_DATE NaN NaN NaN FIRST_FLIGHT_DATE NaN NaN NaN GENDER NaN NaN NaN FFP_TIER 4 4 6 WORK_CITY NaN NaN NaN WORK_PROVINCE NaN NaN NaN WORK_COUNTRY NaN NaN NaN AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0	10111c_1001115110	02000	rari	11011	riari	2. 12010	1.00110
MEMBER_NO 1 31494. 5 62988 FFP_DATE NaN NaN NaN FIRST_FLIGHT_DATE NaN NaN NaN GENDER NaN NaN NaN FFP_TIER 4 4 6 WORK_CITY NaN NaN NaN WORK_PROVINCE NaN NaN NaN WORK_COUNTRY NaN NaN NaN AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0		min	50%	6 max			
FFP_DATE NaN NaN NaN FIRST_FLIGHT_DATE NaN NaN NaN GENDER NaN NaN NaN FFP_TIER 4 4 6 WORK_CITY NaN NaN NaN WORK_PROVINCE NaN NaN NaN WORK_COUNTRY NaN NaN NaN AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0	MEMBER NO						
FIRST_FLIGHT_DATE NaN NaN NaN GENDER NaN NaN NaN FFP_TIER 4 4 6 WORK_CITY NaN NaN NaN WORK_PROVINCE NaN NaN NaN WORK_COUNTRY NaN NaN NaN AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0	_						
GENDER NaN NaN NaN FFP_TIER 4 4 6 WORK_CITY NaN NaN NaN WORK_PROVINCE NaN NaN NaN WORK_COUNTRY NaN NaN NaN AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0	_						
FFP_TIER 4 4 6 WORK_CITY NaN NaN NaN WORK_PROVINCE NaN NaN NaN WORK_COUNTRY NaN NaN NaN AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0	-						
WORK_CITY NaN NaN NaN WORK_PROVINCE NaN NaN NaN WORK_COUNTRY NaN NaN NaN AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0							
WORK_PROVINCE NaN NaN NaN WORK_COUNTRY NaN NaN NaN AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0	——————————————————————————————————————						
WORK_COUNTRY NaN NaN NaN AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0							
AGE 6 41 110 LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0	_						
LOAD_TIME NaN NaN NaN FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0	_						
FLIGHT_COUNT 2 7 213 BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0							
BP_SUM 0 5700 505308 EP_SUM_YR_1 0 0 0	_						
EP_SUM_YR_1 0 0 0	_						
	——————————————————————————————————————						

SUM_YR_1	0	2800	239560
SUM_YR_2	0	2773	234188
SEG_KM_SUM	368	9994	580717
WEIGHTED_SEG_KM	0	6978. 26	558440
LAST_FLIGHT_DATE	NaN	NaN	NaN
AVG_FLIGHT_COUNT	0.25	0.875	26.625
AVG_BP_SUM	0	752. 375	63163.5
BEGIN_TO_FIRST	0	50	729
LAST_TO_END	1	108	731
AVG_INTERVAL	0	44.6667	728
MAX_INTERVAL	0	143	728
ADD_POINTS_SUM_YR_1	0	0	600000
ADD_POINTS_SUM_YR_2	0	0	728282
EXCHANGE_COUNT	0	0	46
avg_discount	0	0.711856	1.5
P1Y_Flight_Count	0	3	118
L1Y_Flight_Count	0	3	111
P1Y_BP_SUM	0	2692	246197
L1Y_BP_SUM	0	2547	259111
EP_SUM	0	0	74460
ADD_Point_SUM	0	0	984938
Eli_Add_Point_Sum	0	0	984938
L1Y_ELi_Add_Points	0	0	728282
Points_Sum	0	6328.5	985572
L1Y_Points_Sum	0	2860. 5	728282
Ration_L1Y_Flight_Count	0	0. 5	1
Ration_P1Y_Flight_Count	0	0. 5	1
Ration_P1Y_BPS	0	0. 514252	0.999989
Ration_L1Y_BPS	0	0.476747	0.999993
Point_NotFlight	0	0	140

D:\Anaconda3\lib\site-packages\pandas\compat_optional.py:106: UserWarning: Pandas requires version '0.9.8' or newer of 'xlsxwriter' (version '0.9.6' currently insta lled).

In [6]:

```
# -*- coding:utf-8 -*-
 数据预处理
from __future__ import division
from pandas import DataFrame, Series
import pandas as pd
datafile = 'air_data.csv'#航空公司原始数据,第一行是属性名
data = pd. read_csv(datafile, encoding='utf-8')
# 1> 数据清洗
# 丢弃掉票价为0的记录; 丢弃票价为0、平均折扣不为零、总飞行公里大于0的记录
cleanedfile = 'cleaned.xlsx'
datal = data[data['SUM YR 1'].notnull()*data['SUM YR 2'].notnull()] #票价非空值才保留,去掉空值
#只保留票价非零的,或者平均折扣率与总飞行公里数同时为零的记录
index1 = data1['SUM_YR_1'] != 0
index2 = data1['SUM_YR_2'] != 0
index3 = (data1['SEG_KM_SUM'] == 0) & (data1['avg_discount'] == 0)
data1 = data1[index1 | index2 | index3] #或关系
data1. to_excel(cleanedfile)
data2 = data1[['LOAD_TIME', 'FFP_DATE', 'LAST_TO_END', 'FLIGHT_COUNT', 'SEG_KM_SUM', 'avg_discount']]
data2. to_excel('datadecrese.xlsx')
D:\Anaconda3\lib\site-packages\pandas\core\computation\expressions.py:194: UserWar
```

ning: evaluating in Python space because the '*' operator is not supported by nume xpr for the bool dtype, use '&' instead op=op str, alt op=unsupported[op str]

D:\Anaconda3\lib\site-packages\pandas\compat_optional.py:106: UserWarning: Pandas requires version '0.9.8' or newer of 'xlsxwriter' (version '0.9.6' currently insta 11ed).

```
In [7]:
# 2> 数据规约
import numpy as np
data = pd. read excel ('datadecrese. xlsx')
data['L1'] = pd. to_datetime(data['LOAD_TIME']) - pd. to_datetime(data['FFP_DATE'])# 以纳秒为单位
print(data['L1'].head())
# data['L3'] = data['L1']. astype('int64')/10**10/8640/30 # 此方法假定每个月是30天,这方法不准确
data['L3'] = data['L1']/np.timedelta64(1, 'M') # 将间隔时间转成月份为单位,注意,此处必须加一个中
print(data['L3'].head())
   2706 days
0
   2597 days
1
2
   2615 days
3
   2047 days
   1816 days
4
Name: L1, dtype: timedelta64[ns]
    88.905316
0
    85.324134
1
2
    85. 915522
3
    67. 253948
    59.664469
Name: L3, dtype: float64
```

In [8]:

```
# 将表中的浮点类型保留至小数点后四为
\# f = 1 \text{ambda x:'} \%. 2f' \% x
# data[['L3']] = data[['L3']].applymap(f) # or data['L3'] = data['L3'].apply(f)
# data[['L3']] = data[['L3']].astype('float64')#注意:使用apply或applymap后,数据类型变成Object,
data["L3"] = data["L3"]. round(2) # 等价于上面三句话,数据类型不变
data['LAST_TO_END'] = (data['LAST_TO_END']/30).round(2)
data['avg discount'] = data['avg discount'].round(2)
data.drop('L1', axis=1, inplace =True) # 删除中间变量
data.drop(data.columns[:3], axis=1, inplace =True) # 去掉不需要的u'LOAD TIME', u'FFP DATE'
data.rename(columns={'LAST_TO_END':'R', 'FLIGHT_COUNT':'F', 'SEG_KM_SUM':'M', 'avg_discount':'C', 'L3'
print(data.head())
data.to_excel('sxgz.xlsx',index=False)
def f(x):
   return Series([x.min(), x.max()], index=['min', 'max'])
d = data.apply(f)
d. to_excel('summary_data.xlsx')
```

```
R
                         C
                                L
           F
                   M
              580717
0
  0.03
        210
                     0.96 88.91
                            85.32
  0.23
        140
              293678
                     1. 25
  0.37
         135
              283712
                      1.25
                            85.92
3
  3. 23
         23
              281336
                      1.09
                            67.25
        152
  0.17
              309928 0.97 59.66
```

D:\Anaconda3\lib\site-packages\pandas\compat_optional.py:106: UserWarning: Pandas requires version '0.9.8' or newer of 'xlsxwriter' (version '0.9.6' currently insta 11ed).

In [9]:

```
# 3> 数据标准化

#标准差标准化

d1 = pd. read_excel('sxgz.xlsx')

d1=d1.astype('float64')

d2 = (d1-d1.mean())/d1.std()

d1 =d2.iloc[:,[4,0,1,2,3]]

d1.columns = ['Z'+i for i in d1.columns]#表头重命名

d1.to_excel('sjbzh.xlsx',index=False)
```

D:\Anaconda3\lib\site-packages\pandas\compat_optional.py:106: UserWarning: Pandas requires version '0.9.8' or newer of 'xlsxwriter' (version '0.9.6' currently insta lled).

warnings.warn(msg, UserWarning)

In [1]:

```
# -*- coding:utf-8 -*-
#模型构建
#使用K-means聚类算法分类并分析每类的特征
import pandas as pd
from pandas import DataFrame, Series
from sklearn.cluster import KMeans #导入K均值聚类算法
k = 5 # 聚为5类
d3 = pd. read excel('sjbzh.xlsx')
#调用k-means算法,进行聚类分析
kmodel = KMeans(n_clusters=k, n_jobs=4)# n_job是并行数, 一般等于CPU数较好
kmodel.fit(d3)
labels = kmodel.labels #查看各样本类别
demo = DataFrame(labels, columns=['numbers'])
demol= DataFrame(kmodel.cluster_centers_, columns=d3.columns) # 保存聚类中心
demo2= demo['numbers'].value_counts() # 确定各个类的数目
demo4 = pd. concat ([demo2, demo1], axis=1)
demo4. index. name='labels'
demo4. to excel ('kmeansresults.xlsx')
```

D:\Anaconda3\lib\site-packages\sklearn\cluster\kmeans.py:939: FutureWarning: 'n_j obs' was deprecated in version 0.23 and will be removed in 0.25.

" removed in 0.25.", FutureWarning)

D:\Anaconda3\lib\site-packages\pandas\compat_optional.py:106: UserWarning: Pandas requires version '0.9.8' or newer of 'xlsxwriter' (version '0.9.6' currently insta lled).

```
In [18]:
```

```
demo2= demo['numbers'].value_counts() # 确定各个类的数目 demo2
```

Out[18]:

- 0 24353
- 4 15657
- 1 12016
- 3 5338
- 2 4680

Name: numbers, dtype: int64

In [19]:

```
print (kmodel.cluster_centers_)#查看聚类中心
print (kmodel.labels_)#查看各样本类别
```

```
 \left[ \left[ -0.70064779 \right. \left. -0.41690258 \right. \left. -0.15788326 \right. \left. -0.15669777 \right. \left. -0.27310712 \right] \right.
```

 $\begin{bmatrix} -0.31019292 & 1.69228907 & -0.57466037 & -0.53664874 & -0.18763601 \end{bmatrix}$

 $\begin{bmatrix} -0.\ 00705242 & 0.\ 00427526 & -0.\ 25033534 & -0.\ 26075461 & 2.\ 06290632 \end{bmatrix}$

[0.48567139 -0.79989636 2.48350667 2.424729 0.31483049]

[1.16377895 -0.37807174 -0.08540491 -0.09321834 -0.15969163]]

[3 3 3 ... 0 1 1]

In []: